Study Title

Weight-of-the evidence evaluation of 2,4-D potential for interactions with the estrogen, androgen and thyroid pathways and steroidogenesis

Data Requirement
Non-Data Requirement Report

Authors
B.H. Neal, J. Bus, M.S. Marty, K. Coady, A. Williams, J. Staveley and J.C. Lamb

Study Completed on:
December 11, 2016

Performing Laboratory
Exponent, Inc.
Alexandria VA
and
Toxicology and Environmental Research and Consulting
The Dow Chemical Company
Midland MI

Published by:
Critical Reviews in Toxicology

Publication Reference ID
Critical Reviews in Toxicology
10.1080/10408444.2016.1272094

Critical Reviews in Toxicology
10.1080/10408444.2016.1272094

STATEMENT OF NO DATA CONFIDENTIALITY CLAIMS

TITLE: Weight-of-the-evidence evaluation of 2,4-D potential for interactions with the estrogen, androgen and thyroid pathways and steroidogenesis

No claim of confidentiality is made for any information contained in this study on the basis of its failing within the scope of FIFRA Section 10(d)(1)(A),(B) or (C).

Company: Industry Task Force II on 2,4-D Research Data

Company Agent: Steve A. McMaster

Title: Technical Director

Signature: Steve A. McMaster

Date: 24 March 2017

STATEMENT OF COMPLIANCE WITH GOOD LABORATORY PRACTICE STANDARDS

This study is not subject to Good Laboratory Practice Standards 40 CFR Part 160.

Submitter:

Company: Industry Task Force II on 2,4-D Research Data

Company Agent: Steve A. McMaster

Title: Technical Director

Signature: Steve A. McMaster

Date: 24 March 2017

Taylor & Francis

Critical Reviews in Toxicology

ISSN: 1040-8444 (Print) 1547-6898 (Online) Journal homepage: http://www.tandfonline.com/loi/itxc20

Weight-of-the-evidence evaluation of 2,4-D potential for interactions with the estrogen, androgen and thyroid pathways and steroidogenesis

B. H. Neal, J. Bus, M. S. Marty, K. Coady, A. Williams, J. Staveley & J. C. Lamb

To cite this article: B. H. Neal, J. Bus, M. S. Marty, K. Coady, A. Williams, J. Staveley & J. C. Lamb (2017): Weight-of-the-evidence evaluation of 2,4-D potential for interactions with the estrogen, androgen and thyroid pathways and steroidogenesis, Critical Reviews in Toxicology

To link to this article: http://dx.doi.org/10.1080/10408444.2016.1272094

CRITICAL REVIEWS IN TOXICOLOGY, 2017
http://dx.doi.org/10.1080/10408444.2016.1272094

Taylor & Francis
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS

Weight-of-the-evidence evaluation of 2,4-D potential for interactions with the estrogen, androgen and thyroid pathways and steroidogenesis

B. H. Neal[a], J. Bus[a], M. S. Marty[b], K. Coady[b], A. Williams[a], J. Staveley[a] and J. C. Lamb[a]

[a]Exponent®, Alexandria, VA, USA; [b]Toxicology & Environmental Research and Consulting, The Dow Chemical Company, Midland, MI, USA

ABSTRACT

A comprehensive weight-of-the-evidence evaluation of 2,4-dichlorophenoxyacetic acid (2,4-D) was conducted for potential interactions with the estrogen, androgen and thyroid pathways and with steroidogenesis. This assessment was based on an extensive database of high quality *in vitro*, *in vivo* ecotoxicological and *in vivo* mammalian toxicological studies. Epidemiological studies were also considered. Toxicokinetic data provided the basis for determining rational cutoffs above which exposures were considered irrelevant to humans based on exceeding thresholds for saturation of renal clearance (TSRC); extensive human exposure and biomonitoring data support that these boundaries far exceed human exposures and provide ample margins of exposure. 2,4-D showed no evidence of interacting with the estrogen or androgen pathways. 2,4-D interacts with the thyroid axis in rats through displacement of thyroxine from plasma binding sites only at high doses exceeding the TSRC in mammals. 2,4-D effects on steroidogenesis parameters are likely related to high-dose specific systemic toxicity at doses exceeding the TSRC and are not likely to be endocrine mediated. No studies, including high quality studies in the published literature, predict significant endocrine-related toxicity or functional decrements in any species at environmentally relevant concentrations, or, in mammals, at doses below the TSRC that are relevant for human hazard and risk assessment. Overall, there is no basis for concern regarding potential interactions of 2,4-D with endocrine pathways or axes (estrogen, androgen, steroidogenesis or thyroid), and thus 2,4-D is unlikely to pose a threat from endocrine disruption to wildlife or humans under conditions of real-world exposures.

ARTICLE HISTORY
Received 17 September 2015
Revised 14 November 2016
Accepted 11 December 2016

KEYWORDS
2,4-Dichlorophenoxyacetic acid; endocrine; toxicology; ecotoxicology; *in vitro*; toxicokinetics; exposure; reproduction

Table of contents

CONTACT James S. Bus jbus@exponent.com Exponent®, 1800 Diagonal Rd. Suite 500, Alexandria, VA 22314, USA

Introduction

2,4-Dichlorophenoxyacetic acid (2,4 D) was the first synthetic herbicide introduced into commerce in 1947, and as a result, 2,4-D is also among the best understood and most thoroughly researched herbicides in the world. Many studies have been conducted on 2,4-D including *in vitro*, ecotoxicological and mammalian toxicological studies. The influence of dose-dependent 2,4-D toxicokinetics (TK) as an important determinant for expression of toxicity has also been characterized in multiple species, including humans. 2,4-D has been the subject of several epidemiological studies and comprehensive reviews as well as multiple studies to characterize potential human and environmental exposure levels (reviewed in Munro et al. 1992; USDA 1998; Garabrant & Philbert 2002; US EPA 2005; Bus & Hammond 2007; Aylward et al. 2010; Burns & Swaen 2012).

2,4-D is one of the most widely used herbicides worldwide and the third most widely used herbicide in the USA and Canada. Its major uses in agriculture are on wheat and small grains, sorghum, corn, rice, sugar cane, low-till soybeans, rangeland and pasture. It is also used on rights-of-way, roadsides, non-crop areas, forestry, lawn and turf and aquatic weeds. Despite its widespread use, urinary levels of 2,4-D are largely undetectable in the general population in Canada and in the USA (CDC 2005; Health Canada 2010).

In recent years, increased attention has been paid to the potential endocrine-modulating effects of environmental and occupational exposures to pesticides and chemicals. Based on its production volume, 2,4-D was recently screened using the US Environmental Protection Agency (US EPA) Tier I Endocrine Disruptor Screening Program (EDSP) assays for potential interactions with the estrogen, androgen, thyroid (EAT) pathways or with steroidogenesis (Coady et al. 2013, 2014). The endocrine activity of 2,4-D also has been comprehensively evaluated in a state-of-art apical *in vivo* extended one-generation reproduction study (EOGRT; Marty et al. 2013). Overall, the toxicological, epidemiological, exposure and biomonitoring information available for 2,4-D represents a useful and comprehensive dataset to further assess the potential for endocrine interactions. Consideration of exposure information is critical for assessing the human relevance of effects restricted to high *in vitro* concentrations or high *in vivo* dosing, and in determining the likelihood and appropriate level of concern for any given effect.

This assessment of potential endocrine pathway interactions reviews the relevant toxicological and ecotoxicological databases including both regulatory toxicological studies (studies required or developed to support US registration) and published literature. The objective of this review was to construct a weight of the evidence (WoE) assessment of potential endocrine pathway interactions and implications for adverse effects on human health and environmental species, with a particular focus on potential for interactions with EAT pathways and with steroidogenesis.

The WoE approach included

- identifying and reviewing studies of 2,4-D conducted for regulatory purposes and selection of studies with endpoints potentially relevant to EAT and steroidogenesis;
- identifying and reviewing published *in vitro*, *in vivo* ecotoxicological, *in vivo* mammalian, epidemiological and mechanistic studies of 2,4-D with endpoints potentially relevant to EAT and steroidogenesis;
- evaluating study quality for both regulatory and published studies;
- identifying relevant endpoints for each potential endocrine pathway interaction (EAT or steroidogenesis) within each study;
- ranking endpoints for specificity and sensitivity in the context of the specific studies;
- identifying potential confounding factors;
- evaluating the consistency and coherence of the reported findings suggesting potential pathway interactions, i.e. testing the hypothesis that the compound may act as an estrogen agonist or antagonist, an androgen agonist or antagonist, a thyroid agonist or antagonist or a modulator of steroidogenesis;
- assessing the completeness of the available data and
- developing conclusions on the likelihood of compound-related impacts for each potential EAT or steroidogenesis endocrine pathway interaction.

Selection of regulatory toxicology studies for inclusion in the WoE

Toxicology studies of pesticides in the USA conducted for regulatory purposes (referred to as "regulatory toxicology

studies") are conducted to support product registration. The goal of these studies, many of which are required by the US EPA under contemporaneous detailed study guidelines, is to develop adequate data so that the US EPA can have high confidence in the toxicological endpoints serving as the basis for its short, intermediate or long-term risk assessments to protect worker health, other potentially exposed populations, e.g. exposed through ingestion of dietary residues and the environment. In addition to guideline compliance, the studies are required to be conducted under Good Laboratory Practice (GLP) regulations.

The EDSP Tier 1 testing performed for 2,4-D included five mechanistic *in vitro* assays: estrogen receptor (ER) binding and transactivation, androgen receptor (AR) binding, aromatase inhibition and steroidogenesis. Additionally, an amphibian metamorphosis assay (AMA), which focuses on potential thyroid effects, and a fish short-term reproductive assay (FSTRA), which has endpoints sensitive to estrogen and androgen pathway interactions, were performed for 2,4-D. At the time the Tier 1 EDSP requirements were promulgated, 2,4-D had already been tested in mammals in an EDSP Tier 2 equivalent EOGRT study. This study had multiple endocrine system-related endpoints that were specifically added in consultation with the US EPA and the Pest Management Regulatory Agency of Health Canada (PMRA). This battery of EDSP studies constitute a robust core body of information for the WoE evaluation of the potential of 2,4-D for EAT and steroidogenesis pathway interactions.

Fourteen other regulatory studies with the most relevant and comprehensive endocrine pathway-related endpoints were selected for this WoE evaluation. The majority of the selected regulatory studies were mammalian toxicology studies, including: an EPA guideline two-generation reproductive toxicity study, developmental toxicity studies in rats and rabbits, subchronic toxicity studies in rats, mice and dogs and chronic toxicity studies in rats, mice and dogs. This extensive mammalian regulatory toxicological data-base provides an opportunity to evaluate consistency of responses across species and strains, and also across exposure durations. A single relevant ecotoxicological study, a one-generation quail reproductive toxicity study, was identified. Other ecotoxicological regulatory studies included too few relevant endpoints to be useful in an endocrine WoE evaluation.

Many of the regulatory toxicity studies of 2,4-D (including EDSP studies) have been published. We cite the publications as well as the laboratory study reports. For all regulatory studies with 2,4-D, the laboratory study reports were used to evaluate study quality and derive results to include in the WoE. The reports provide methodological details including protocol, amendments and protocol deviations, supporting data on compound identity, purity and dose confirmation analyses and comprehensive results, including both summary and individual animal results, generally not available in published articles.

The guidelines for regulatory toxicology studies have undergone significant changes over time. Many parameters have been added which more completely characterize potential endocrine pathway-related effects. Thus more recent regulatory studies were prioritized for inclusion in the WoE while some older studies were omitted in that the results of these studies are largely supplanted by the findings derived from higher quality and more comprehensive protocols.

Literature search and selection of published studies for inclusion in the WoE evaluation

Databases searched included: BHCAPLUS, MEDLINE, AGRICOLA, CABA, BIOSIS, ESBIOBASE, EMBASE, TOXCENTER, PASCAL, PQSCITECH and SCISEARCH. The search included studies conducted on all forms of 2,4-D including the acid, salt and ester forms. Terms used in the search strategy are included in Supplementary Appendix I. This search covered studies published from 2009 to mid-2013 (depending on the database); further PubMed searches were done to ensure capture of relevant late 2013–early 2014 relevant studies. A similar literature search was conducted in 2009, and coverage of literature extended to the early 1960s in PubMed. Retrieval was limited to English language articles.

Multiple *in vitro*, ecotoxicological and mammalian toxicological studies on 2,4-D were identified from the published literature with potentially relevant endpoints.

Evaluation of study quality

All studies were evaluated for quality prior to inclusion in the assessment, and perceived weaknesses or gaps in available information were tabulated. Although regulatory toxicology studies are typically conducted under GLP, which ensures an *a priori* protocol, a record of any protocol amendments and deviations, and accuracy of data collection and reporting, GLP compliance alone does not guarantee that the studies are of high scientific quality or are the most relevant for evaluation of endocrine pathway modulation. A detailed evaluation of both the regulatory and published studies cited in this article was conducted, and study deficiencies identified where they were found.

Modified Klimisch criteria (Klimisch et al. 1997) were used for scoring study quality and included several additional

Table 1. Modified Klimisch criteria.

Score	Description	Comments
1	*Reliable without restriction*	Generally guideline and GLP compliant studies using validated methodology; study is transparently reported and report internally consistent.
2	*Reliable with restriction*	Generally studies from the published literature, often not GLP compliant but sufficient to accept the data and "scientifically acceptable," or GLP studies that do not follow a specific guideline or cover limited components of a guideline.
3	*Not reliable*	Studies generated by a method that is not acceptable, insufficiently documented, or not convincing using expert judgment.
4	*Not assignable*	Studies reported only in abstracts or as part of book chapters with insufficient detail to evaluate study quality.

factors suggested by Schneider et al. 2009, in their discussion of a Tox-R Tool for study quality evaluation. The scores are summarized in Table 1.

Study quality factors evaluated in *in vitro* studies are

- identification of compound name and purity;
- description of test method;
- identification of dose concentrations;
- rationale (if any) for dose selection;
- use of positive control;
- use of appropriate vehicle or solvent control;
- cytotoxicity evaluations (when appropriate);
- internal consistency and, when tested, reproducibility of reported results;
- biological plausibility of reported results and
- compliance with regulatory guideline or otherwise validated and scientifically appropriate methodology.

There was an exception to application of Klimisch scoring to *in vitro* assays for the assays from EPA's *in vitro* ToxCast™ Program (US EPA 2010). There is insufficient information available on the methodology of the majority of the proprietary ToxCast™ assays to develop a Klimisch score; typically, these would be scored "4" for lack of information. However, data derived from ToxCast™ include assays specifically designed to elucidate potential endocrine-active mechanisms, and currently are being considered by EPA for use in priority setting for the screening of chemicals under EDSP. The selected studies from this program used for the 2,4-D WoE include those most similar in design to the EDSP Tier 1 *in vitro* assays and are therefore considered relevant to the WoE.

Study quality factors evaluated for *in vivo* ecotoxicological and mammalian studies are

- identification of compound and purity;
- description of test method;
- identification of dose concentrations;
- rationale (if any) for dose selection;
- adequacy of method or limitations;
- parameters evaluated and methods used for evaluation;
- completeness of data including identification of source, age and strain of test species;
- number of animals and dose groups tested;
- use of appropriate statistical methods;
- information on analytical dose confirmation, homogeneity and stability of dosing formulations;
- appropriate randomization procedures including accounting for potential litter effects in developmental, reproductive or perinatal studies;
- internal consistency of reported results;
- presence or absence of dose response;
- biological plausibility of reported results and
- compliance with regulatory guideline or otherwise validated and scientifically appropriate methodology.

In general, studies with a Klimisch criteria score of 1 or 2 are included in the WoE; however all studies were reviewed for potentially relevant information.

No attempt was made to score epidemiological or occupational health studies; study limitations are generally discussed. Additionally, non-guideline mechanistic studies were not scored because these studies often use single dose levels and/or unconventional routes of exposure. Klimisch et al. (1997) assign this type of study a "5", outside of the scoring criteria.

Published mammalian toxicological studies of 2,4-D salts and esters (Charles et al. 1996a, 1996b, 2001) were the primary source of information regarding activity of the salts and esters and were reviewed to determine only whether any of these forms presented a unique hazard of endocrine-related toxicity compared to the acid form; therefore, the original study reports were not reviewed in depth.

Identification of relevant endpoints for each potential endocrine pathway interaction and ranking of endpoints for sensitivity and specificity

Each study design was examined to determine endpoints potentially relevant to specific endocrine pathway interactions. The endpoints selected will be reviewed in the WoE discussion. For mammals, endpoints include: developmental landmarks (anogenital distance (AGD), nipple retention in males, vaginal opening and balano-preputial separation); estrous cyclicity; reproductive organ weight and histopathology; mammary gland histopathology; sperm parameters; ovarian follicular counts; thyroid hormones, weight and histopathology; adrenal weight and histopathology; and pituitary weight and histopathology. The EOGRT and two-generation reproductive toxicity studies provide the majority of relevant mammalian endpoints, particularly in the absence of the EDSP Tier 1 mammalian screening studies which were not required for 2,4-D because of the availability of the comprehensive EDSP Tier 2-equivalent EOGRT study. Further, we consider the most robust data to be derived from studies which have internal checks for consistency because of evaluation of similar endpoints across life stages. For example, there were sporadic findings of testicular atrophy in the EOGRT study parental generation. These findings were of low incidence and within historical control range, but more importantly, were not seen in the F1 generation adults, even after a longer duration of exposure to 2,4-D. (Study results were also examined to confirm there were no increases in implantation loss or fetal deaths that could have signified potential culling of a sensitive sub-population.) Thus, the F1-generation results provide additional confidence that the findings in the parental generation were not exposure related.

Subchronic and chronic toxicity studies, however, often have information on reproductive organ weight and histopathology, and sometimes hormone data (e.g. thyroid hormones T4 and thyroid-stimulating hormone (TSH)). Oncogenicity studies may shed light on potential endocrine-mediated toxicity by increases or decreases in certain tumor types. Subchronic and chronic studies may also help identify differences due to route of administration or varied responses due to species or strain differences. Therefore, these studies were also included in the WoE.

As noted previously, the AMA focuses primarily on thyroid-related endpoints as development of the tadpole is highly dependent on thyroid hormone economy. The FSTRA and the quail one-generation study provide additional apical ecotoxicological studies, providing information on potential estrogen, androgen or steroidogenesis pathway interactions. Endpoints in the FSTRA include vitellogenin (VTG) measurements, presence or absence of nuptial tubercles and gonadal histopathology; in the quail, endpoints include fertilization, eggshell thickness and hatching.

In a recent paper, Borgert et al. 2014 proposed the following ranking scheme for evaluating the endpoints assessed in the EDSP Tier 1 Tests for relevance, sensitivity and specificity to testing hypothesized endocrine pathway interactions:

> "Rank 1 was assigned to *in vivo* endpoints that characterize the fundamental physiological actions for androgen, estrogen and thyroid activities. Rank 1 endpoints are specific and sensitive for the hypothesis, interpretable without ancillary data, and rarely confounded by artifacts or non-specific activity. Rank 2 endpoints are specific and interpretable for the hypothesis but less informative than Rank 1, often due to oversensitivity, inclusion of narrowly context-dependent components of the hormonal system (e.g. *in vitro* endpoints) or confounding by non-specific activity. Rank 3 endpoints are relevant for the hypothesis but only corroborative of Ranks 1 and 2 endpoints."

Note that these rankings are made for each relevant end point, not for each assay as a whole. Ranking of endpoints is preset depending on the assay type and relevance to the hypothesis being tested, i.e. whether estrogenicity, anti-estrogenicity, androgenicity, anti-androgenicity or impact on steroidogenesis or on the hypothalamic–pituitary–thyroid (HPT) axis. It should be noted that there is currently, to our knowledge, no agreed-upon quantitative weighting system for specific potentially endocrine-related parameters for studies outside of the EDSP screening studies.

Three other factors were considered when scoring the individual assay parameters. The first is the context of the parameter and how or if potential confounding factors are eliminated or controlled. For example, a relatively high degree of confidence for assessing potential estrogenicity can be placed on uterine weights in uterotrophic assay study animals, which are either ovariectomized and hence not cycling or for uterine weights in immature females. Uterine weights in reproductive toxicity study or subchronic toxicity study females, if they are cycling, however, are not reliable endpoints because the uterine weight varies markedly with the stage of the estrous cycle at the time of necropsy (Stoker & Zorrilla 2010). In the latter case, higher confidence in a potential endocrine interaction would be made if other correlating endpoints, particularly in the same study, also showed a response suggesting estrogenic activity. For example, if uterine weights were increased, *and* if the females showed persistent estrus, this would provide a much more robust signal of potential estrogenic activity.

Second, the magnitude of responses needs to be evaluated carefully. In the context of evaluating potential estrogenicity in rat pubertal development, slight advancement of the time of vaginal opening, e.g. 0.5 day, is not strong evidence of potential estrogenicity, whereas a three-day advance would be (Edwards & Kay 1985). Evaluation of response magnitude requires an appreciation of the variability inherent to the parameter in control test systems or species. Historical control data (HCD) are particularly useful in evaluating whether a statistically significant change is also biologically significant. One of the strengths of the regulatory database generally lacking in other studies in the published literature (and still being developed for many relatively newer endpoints in the EDSP data set, particularly in fish and frogs) is the availability of HCD to help interpret the biological significance of responses, and to determine if the control population is behaving normally.

The third factor that may influence the scoring is the presence of significant systemic toxicity that may confound the ability to accurately ascribe changes to endocrine modulation. For example, decreases in VTG levels in female fish may be due to other toxicity, such as hepatic toxicity, rather than to potential anti-estrogenicity, whereas a substantial increase in VTG in male fish appears to be closely associated with estrogenicity. Decreased weight gain or weight loss may lead to a decreased incidence of mammary tumors or cell proliferation in chronic studies, delays in sexual maturation, and, particularly in immature animals, decreased testis weight and testicular atrophy.

Differentiating potential endocrine modes of action based on endpoints affected

There is overlap between changes in endpoints that may be relevant for either estrogenic or anti-androgenic modes of action. For example, relatively potent estrogens may affect testicular histopathology in ways congruent with anti-androgens. Interactions with ERs or ARs may help delineate modes of action. Endpoints relevant to steroidogenesis or the hypothalamic–pituitary–gonadal (HPG) axis may overlap with either of these mechanisms. Examples of potential indicators of endocrine pathway interactions in mammalian systems are shown in Table 2.

To limit extensive redundancy in the WoE of potential estrogen pathway interactions in mammals we have taken the approach of limiting tabulated endpoints and discussions to female-specific endpoints; for evaluation of potential androgen pathway interactions we focus on male-specific endpoints. Endpoints from the opposite sex are potentially relevant (as can be seen in Table 2) and will be mentioned in each case but not discussed in detail. For fish, the most sensitive indicators of potential endocrine modulation in the current EDSP Tier 1 assay appear to be found in the opposite sex, e.g. increased VTG in male fish is a sensitive endpoint for estrogenicity and the appearance of nuptial tubercles in female fathead minnows is sensitive for androgenicity. Therefore, data from both sexes from the FSTRA are considered for each pathway hypothesis.

Changes in both male and female endpoints may also be indicators of a potential interaction with steroidogenesis or the HPG axis; for this evaluation we have tabulated relevant endpoints in both sexes for both mammals and fish, and provided briefer discussions of any specific study endpoints already discussed in the WoE for estrogen-pathway-related or androgen-pathway-related endpoints.

Table 2. Examples of endpoints in mammalian toxicological studies indicating potential interactions with the estrogen or androgen pathways.

Mechanism	Potential effects in males	Potential effects in females
Estrogenicity	• Delayed preputial separation (marked, or in absence of significant body weight decreases) • ↓ sperm counts • ↓ fertility • Presence of hypospadias/epispadias • ↓ reproductive organ weight (particularly prostate, seminal vesicles) • Histopathological findings of the testes (e.g. leydig cell proliferation and/or tumors) • ↑ incidence, growth of pituitary tumors	• Accelerated vaginal opening • Persistent estrus • ↑ time to mating • ↓ gestation duration • ↑ uterine weights (particularly indicative in immature or ovariectomized animals) • Histopathological findings of the female reproductive organs (e.g. vaginal cornification, uterine hypertrophy and hyperplasia) • ↑ mammary tumors • ↑ incidence, growth of pituitary tumors • Ovarian, uterine and vaginal tumors in female offspring
Anti-estrogenicity	• ↓ Height of epithelium in testicular tubules • ↑ testicular weight (short term) • Testicular atrophy (long term) • Infertility and testicular atrophy (moderate term)	• Delayed vaginal opening (VO) • Delayed start of estrous cycling • Irregular or absent estrous cyclicity • ↓ fertility • ↓ corpora lutea, implantations • ↓ female reproductive organ weights • ↓ mammary tumor incidence
Androgenicity	• ↑ or ↓ male reproductive organ weights • ↓ sperm counts • Testicular atrophy	• ↑ Anogenital distance • Accelerated vaginal opening • ↓ fertility • Altered differential follicle count • Histopathological findings of the female reproductive organs (e.g. vaginal agenesis) • Induced male sex accessory tissues
Anti-androgenicity	• Delayed preputial separation • ↓ anogenital distance • Ectopic testes (pre-natal exposure) • Hypospadias/epispadias (pre-natal exposure) • ↓ fertility • ↓ reproductive organ weight (particularly prostate, seminal vesicles) • Retained nipples/areolas in male pups • Histopathological findings of the reproductive organs (e.g. epididymal agenesis, testicular tumors)	• Altered pup sex ratios between external and internal sexing
Aromatase inhibition	• ↑ time to mating • ↓ male mounting behavior • ↓ body weight (chronic) • ↑ testis weight (chronic)	• ↑ Body weight • ↓ or ↑uterine weight • ↑ ovary size • Polycystic ovaries • Stromal hyperplasia in ovary (chronic) • Hyalinization in ovary (chronic) • ↑ ureter and bladder infection • ↓ mammary tumor incidence (S-D rats)
Reduced steroid biosynthesis	• Similar to anti-androgenicity • Possible increased serum cholesterol levels • ↑ incidence Leydig cell tumors	• Similar to anti-estrogenicity/aromatase inhibition

↑ Increase
↓ Decrease

The WoE generally follows the approach used by de Peyster and Mihaich (2014), in that potentially relevant studies are identified, studies are evaluated for quality, relevant endpoints for each hypothesis tested were ranked for sensitivity and specificity, other factors or confounders potentially influencing each endpoint were evaluated, and, most importantly, the consistency of responses of relevant endpoints is assessed. The goal in this evaluation has been to use the most transparent methods for evaluation possible, recognizing that as more information on adverse outcome pathways are developed, some of the relative rankings of endpoints may change accordingly.

The Weight of Evidence Guidance for the Tier 1 EDSP studies developed by the US EPA (US EPA 2011) was also considered when developing this evaluation. This document indicates:

> "The robustness of the Tier 1 battery is based on the strengths of each individual assay and the complementary endpoints within the battery. Thus, "*... the value of each individual assay cannot be considered in isolation from other assays in the battery, as they have been combined in a manner such that limitations of one assay are complemented by the strengths of another*"" (quote in EPA document from EDSTAC, 1998).

Although EPA's approach was developed specifically for the Tier 1 EDSP data set, the same principles were followed for the current evaluation of 2,4-D. The WoE reflects an assessment of whether results might signal a specific endocrine pathway interaction, the relative weight or rank placed on that parameter for specifically and sensitively flagging a potential interaction, and whether a finding (if any) was made only at a systemically toxic or otherwise excessive dose, as discussed above. The WoE tables developed for each pathway provide a visual representation that assists in identifying patterns of findings within or across studies that may indicate a potential endocrine pathway interaction; the

subsequent discussions evaluate potentially confounding or other factors that need to be considered to evaluate the likelihood of an endocrine pathway interaction.

Impact of toxicokinetic data for 2,4-D on study design, data interpretation and risk assessment

Extensive research has been done to characterize 2,4-D TK. 2,4-D clearly exhibits species-, dose- and sex-dependent non-linear TK in animal test species (Gorzinski et al. 1987; Van Ravenswaay et al. 2003; Timchalk 2004; Saghir et al. 2006; 2013). The non-linear TK is directly and primarily attributable to high-dose saturation of a renal anion transporter, OAT-1, that is responsible for rapid renal clearance of 2,4-D (Hasegawa et al. 2003; Saghir et al. 2013). Non-linear TK, in which metabolism or excretion pathways available at lower blood concentrations are partly or fully saturated with increasing dose, may be a clear confounder both in appropriately designing toxicological studies and evaluating results for hazard and human risk assessment. Use of TK to inform human-relevant dose selection in animal toxicity studies has been affirmed in recent reviews and Organization for Economic Cooperation and Development (OECD) guidance on conduct of the EOGRT study (Barton et al. 2006; Carmichael et al. 2006; Cooper et al. 2006; OECD 443, 2012a). The guidance recommended that top dose level(s) should not exceed the inflection point of onset of TK non-linearity if the inflection point dose was well separated from human exposures, and further concluded that toxicity limited to doses above the onset of non-linear TK behavior was not quantitatively relevant to human risk. Both of these criteria, evidence of non-linear TK in animal test systems and low human exposure levels, are fulfilled for 2,4-D.

Consideration of saturated TK is particularly important for interpretation of the human health relevance of high-dose specific 2,4-D toxicity, including potential endocrine effects. 2,4-D is a structural analog of thyroxine, and has been shown to be weakly active in displacing plasma protein bound thyroxine following administration at a toxicokinetically saturated 80 mg/kg/day dose in rats (Florsheim & Velcoff 1962; Florsheim et al. 1963; Van den Berg et al. 1991). Given the weak competitive binding activity of 2,4-D to thyroxine binding sites, any substantial displacement would be unlikely at disproportionately lower plasma concentrations associated with non-saturating 2,4-D doses. In addition, high-dose administration of 2,4-D to mice and rabbits results in increased distribution to and/or retention in brain (Kim et al. 1988). The overall TK data suggest that potential central nervous system (CNS) effects are secondary to two sequential and mechanistically related dose disproportionate events resulting in increases in 2,4-D brain concentrations: (1) initial saturation of OAT-1 renal clearance leading to dose-disproportionate elevation in plasma 2,4-D plasma concentration allowing for increased organ distribution of non-plasma-protein bound 2,4-D (Timchalk 2004; van Ravenswaay et al. 2003); and (2) followed by augmented non-linear increases in brain concentration associated with high-dose specific saturation of OAT-1 clearance from brain (Kim et al. 1988). Although the quantitative contribution of each of these saturation events to altered distribution of 2,4-D within the brain is unknown, such alterations to and within an endocrine modulatory organ such as brain have the potential to initiate high-dose specific secondary modes of action. These include reduced clearance of potentially toxic neurotransmitter metabolites such as 5-hydroxy-3 indole acetic acid from brain by the choroid plexus OAT-1 transporter (5-HIAA; Kim et al. 1988; Elo & MacDonald 1989) that ultimately have no quantitative relevance to adverse health outcome potential in humans exposed to far lower, non-saturating, environmental exposures.

Studies in rats have confirmed that 2,4-D TK exhibits non-linear behavior following dietary administration, a route of administration commonly employed in 2,4-D toxicity studies, and have titrated the doses at which the inflection point of onset of non-linear TK begins in both male and female rats (Saghir et al. 2008a, 2008b, 2013). In early work using dietary dose levels of 5 and 100 mg/kg/day, Saghir and coworkers demonstrated saturation of renal clearance and distinctly non-linear TK in male F344 rats fed diet approximating 100 mg/kg/day 2,4-D for 28 days (Saghir et al. 2006). However, to better inform dose selection for the EOGRT study, more comprehensive dietary range finding and TK studies were conducted over multiple life stages in both sexes of CD® rats (Saghir et al. 2008a, 2008b, 2013). These data provided information on plasma concentrations over a wide range of 2,4-D dietary doses and identified inflection points for transition from linear to non-linear TK in both male and female Sprague-Dawley rats.

Following an integrated analysis of the TK information, toxicity and human exposure information, top doses of 600 ppm (30 mg/kg/day, non-pregnant females) and 800 ppm (40 mg/kg/day, males) were selected for the EOGRT study. These doses were anticipated to be either at or slightly above the inflection point for non-linear TK behavior of 2,4-D, considered a threshold for saturation of renal clearance (TSRC) for each respective gender (also referred to as a KMD or kinetically derived maximum dose in some reports, Saghir et al. 2012). Data from the EOGRT TK range finder study (Saghir et al. 2013) showed that the high dose for male rats (800 ppm; 41 mg/kg/day) was close to, but slightly above the TSRC. Following 28 days of dietary treatment prior to mating, the 2,4-D plasma area under the curve (AUC) in the 800 ppm dose was a dose-disproportionate 11-fold higher relative to the AUC at the 8-fold lower 100 ppm dose (5 mg/kg/day). The high dose of 600 ppm selected for females in the EOGRT study, however, substantially exceeded the TSRC during the 28-day pre-mating treatment. During the 28-day pre-mating dosing period, the female plasma 2,4-D AUCs at 200, 400 and 600 ppm doses (14, 25–27 and 41 mg/kg/day, respectively) were 3-, 8–11- and 31-fold higher relative to the AUC at 100 ppm (6–7 mg/kg/day). An even larger 33-fold difference in plasma AUC was observed between the 100 and 600 ppm doses on gestation day 17 rat dams (Marty et al. 2013), likely due to increased food consumption in the latter stages of pregnancy. Based on the EOGRT range finder TK data, the TSRC in adult male rats is 30–40 mg/kg/day and 15–20 mg/kg/day in adult non-pregnant females. Males are more efficient at excreting 2,4-D than females because androgens increase the expression of the saturable OAT-1

transporter (Ljubojevic et al. 2004). Thus, the sex-dependent difference in expression of the organic ion transporter likely accounts for the differential thresholds for saturation of 2,4-D between male and female rats.

Differences in species sensitivity to 2,4-D also appear related to the presence or absence of the OAT-1 transporter in the renal tubules. The implications of species-specific differences in 2,4 D TK to selecting the appropriate species for deriving the point of departure (POD) for human risk assessment have been summarized in a review by the Industry Task Force II on 2,4-D Research Data (Bus & Hammond 2007):

> "Knowledge of the dose and species-dependent pharmacokinetic behavior of 2,4-D significantly enhances the understanding of the relevance of toxicity findings of 2,4-D in rodents, and particularly in dogs, to predicting potential human health risks. Once absorbed, 2,4-D is rapidly and completely excreted in urine by both rats and humans, but not dogs (Van Ravenswaay et al. 2003; Timchalk, 2004). In rodents and humans, renal excretion of 2,4-D is facilitated by a saturable organic anion active transporter located in the renal tubules (Timchalk, 2004). The transporter does not effectively function in dogs. Studies in rats indicate the renal clearance of 2,4-D is clearly saturated at oral [gavage] dose levels of 50 mg/kg, resulting in nonlinear increases in 2,4-D blood concentrations at this dose and above (Gorzinski et al. 1987; Van Ravenswaay et al. 2003). Given this non-linear behavior, saturation of 2,4-D renal clearance at 50 mg/kg suggests that animal toxicity findings observed at this dose level and higher overestimate potential human risks. In the case of dogs, both subchronic and chronic studies indicate this species, with an overall NOAEL of 1 mg/kg/day (Charles et al. 1996b), is more sensitive to 2,4-D-induced toxicity than rodents, with an overall NOAEL of 5 mg/kg/day (Charles et al. 1996c). Since the dog is lacking an effective renal organic anion clearance mechanism, this differential species response has been attributed to an inability of the dog to effectively clear 2,4-D from the body, resulting in significantly higher 2,4-D blood concentrations in dog relative to rats and humans at an equivalent oral dose of 5 mg/kg (Van Ravenswaay et al. 2003; Timchalk, 2004). In this case the rat represents a more relevant species for deriving data for [human] risk assessment."

Because of the substantial differences in TK of 2,4-D in dogs relative to other species including humans, EPA, Canadian PMRA and European Food Safety Authority (EFSA) regulatory assessments of 2,4-D have concluded that the dog is an inappropriate species for human risk assessment (US EPA 2005; PMRA 2007; EFSA 2014). As a consequence, the animal no-observed-adverse-effect level (NOAEL) used as the primary reference point to establish acceptable chronic human 2,4-D exposures is 21 mg/kg/day based on toxicity in chronic dietary studies in rats (Marty et al. 2013). This NOAEL is based on renal toxicity, not on endocrine or reproductive effects. Alexander et al. (2007) reported that children living on farms on which 2,4-D was being actively applied had systemic doses (geometric mean) of 0.32 (children 4–11) to 0.12 (children >12 years old) µg/kg based on five days of comprehensive urinary biomonitoring. These dose levels were 65 625–175 000-fold below the overall NOAEL of 21 mg/kg/day (21 000 µg/kg/day) used to set the EPA chronic reference dose for 2,4-D. Large margins of exposure (MOEs) were similarly noted for both applicators and spouses (geometric mean systemic doses of 2.46 and 0.8 µg/kg/day, respectively). Biomonitoring equivalent determinations in this and other populations similarly demonstrate conservatively large MOEs (Aylward et al. 2010; Hays et al. 2012). Since the inflection points for onset of non-linear TK in male and female rats are in the range of 15–40 mg/kg/day, toxicity studies such as the EOGRT fulfilled recent dose selection guidance recommending use of a KMD dose selection strategy, i.e. for 2,4-D using doses at or below the TSRC, when the non-linear TK inflection point is well separated from human exposures. Thus, TK data are a key contextual consideration facilitating interpretation of the potential human relevance of 2,4-D toxicity findings limited to doses above the TSRC, including potentially endocrine-related endpoints.

Organization of the WoE evaluation

The WoE is organized to summarize the 2,4-D *in vitro* studies, followed by studies from the ecotoxicological and mammalian toxicological databases for 2,4-D with endpoints relevant to evaluating EAT and steroidogenesis endpoints. EDSP Tier 1 studies, and the quail one-generation reproductive toxicity study, the EOGRT EDSP Tier 2-equivalent and multi-generation rat studies are summarized briefly in the appropriate sections because these studies provide the most relevant information for characterizing EAT or steroidogenesis interactions. In each case, the available studies from the regulatory databases and the published literature are tabulated with a brief description of method, results, Klimisch score and rationale. The published *in vitro* studies are presented alphabetically by first author, because many of these publications cover multiple types of *in vitro* assays. The ecotoxicological and mammalian toxicological studies are organized by study type. Following the review of mammalian studies is a brief overview of epidemiological studies that assessed relevant endpoints.

The article continues with the WoE assessments for potential interactions with the estrogen, androgen or thyroid pathways or for interaction with steroidogenesis or HPG axis integrating the data from all studies considered high quality (Klimisch 1 or 2).

There are several supplementary appendices. The first, Appendix I, provides the search strategy used in to identify potentially relevant published literature. Six appendices follow that include more comprehensive summaries of the regulatory toxicological studies and published studies:

- Appendix II: *in vitro* studies (Klimisch criteria 1 or 2), including
  - EDSP *in vitro* studies
  - *In vitro* studies in the published literature
  - Further detail on ToxCast™ assays
- Appendix III: *in vivo* ecotoxicological studies (Klimisch criteria 1 or 2), including
  - EDSP *in vivo* ecotoxicological studies
    - Amphibian metamorphosis assay (AMA)
    - Fish short term reproduction (FSTR) assay
  - Quail one-generation reproductive toxicity study
  - *In vivo* ecotoxicological studies in the published literature

- Appendix IV: *in vivo* mammalian toxicological studies (Klimisch criteria 1 or 2), including:
  - Reproductive toxicity
    - EDSP Tier 2 equivalent EOGRT study
    - Guideline two-generation rat reproductive toxicity study
  - Developmental toxicity studies in rat and rabbit
  - Subchronic and chronic toxicity studies in rats, mice and dogs
  - *In vivo* mammalian toxicological studies in the published literature

Appendix III includes summaries of the regulatory ecotoxicological summaries on 2,4-D acid, followed by published ecotoxicological studies. Appendix IV includes summaries of the regulatory mammalian toxicological summaries on 2,4-D acid, followed by published mammalian studies organized by study type, with priority given to the types of studies with the most relevant endpoints for assessing potential endocrine pathway interactions, e.g. reproductive toxicity evaluations.

Studies considered to be of less than optimal quality for inclusion in the WoE (Klimisch 3 or 4), or found to not contain relevant data are summarized in Appendices V (*in vitro* studies); VI (*in vivo*-ecotoxicological studies) and VII (*in vivo*-mammalian studies).

In vitro studies of 2,4-D relevant to assessment of potential endocrine pathway interactions

In general, *in vitro* studies may assist in defining adverse outcome pathways and clarifying *in vivo* findings, but are not indicative by themselves of an adverse endocrine-disrupting effect. Further, results of these studies may be influenced by incompletely or unassessed cytotoxicity, artifacts from transient cell transfection, lack of metabolic co-factors or irrelevant compound concentrations tested.

EDSP tier I in vitro *studies*

The *in vitro* EDSP screening assays of 2,4-D are described in a recent publication (Coady et al. 2014). These assays followed the published US EPA guidelines for ER binding (rat uterine cytosol ER binding assay), ER-mediated transcriptional activation (HeLa-9903-ERα transactivation assay), AR binding (rat prostate cytosol AR binding assay), aromatase enzymatic activity inhibition (recombinant human CYP19 aromatase inhibition assay) and interference with steroidogenesis (H295R steroidogenesis assay).

The single exception to the guidelines for these assays was that it was considered appropriate to limit the high concentration in the first four EDSP *in vitro* assays to 100 µM, rather than the guideline-recommended 1 mM, because the 100 µM concentration was equivalent to serum concentrations at or slightly above the TSRC (slightly above the inflection point for non-linear TK) in rats dosed with 2,4-D in the diet in the EOGRT study (Marty et al. 2010; Marty et al. 2013; Saghir et al. 2013). As noted previously, responses seen only in the non-linear TK range are not regarded as relevant to human risk assessment. Thus, endocrine receptor binding or activation observed only at *in vitro* concentrations equal to or exceeding serum concentrations at the TSRC are not regarded as relevant to human risk and therefore not informative of potential human endocrine risk; testing high-exaggerated concentrations was considered not appropriate or useful. The maximum concentration recommended in the steroidogenesis assay is 100 µM, and was used in that assay.

These EDSP studies are considered to meet Klimisch criteria 1 because they were conducted according to US EPA guideline recommendations, methodology was validated extensively, deviations from the method were minor and performance and reporting complied with GLP. Table 3 below summarizes the assay type, concentration range tested, results and study quality evaluation from the EDSP *in vitro* studies; detailed summaries of methods and results are provided in Supplementary Appendix IIA1. The ER binding, ER transactivation, AR binding and aromatase assays showed no effects of 2,4-D, predicting no interactions with the estrogen or androgen pathways either as an antagonist or agonist. There was no effect on testosterone level in the steroidogenesis assay. There was a statistically significant increase in estradiol at the highest concentration tested in all three replicates of the steroidogenesis assay; however, the magnitude of the change was very low (1.2 fold) and did not meet the 1.5-fold cutoff criterion for an exposure-related increase established in the steroidogenesis assay validation studies (Hecker et al. 2008). Therefore, it was concluded that there was no robust evidence for an exposure-related effect.

Published in vitro *studies*

In addition to the *in vitro* Tier 1 EDSP screening assays described by Coady et al. (2014) and summarized above, sixteen additional published *in vitro* studies investigating the potential endocrine activity of 2,4-D were identified. Many of these publications contain multiple assays. These are listed in Table 4 and include: studies of ER and AR agonist and antagonist activity as measured in transactivation assays, assays of ER, AR and progesterone receptor (PR) binding, tissue steroid hormone production and the proliferation of estrogen-responsive cells.

Studies with Klimisch scores of 1 or 2 are summarized in Supplementary Appendix II B; other studies are summarized in Supplementary Appendix V. Supplementary Appendix V also provides a general explanation for the exclusion of yeast-based assays, although these assays were reviewed.

ToxCast™ assays of 2,4-D

EPA developed the ToxCast™ program as a high throughput *in vitro* screen (HTS) using primarily proprietary assays to screen for potential biological activity and to be used, in conjunction with exposure information, to prioritize chemicals for future testing. EPA has recognized that ToxCast assays offer quantitative data informing the potential reactivity of substances with endocrine pathways and thus can serve as alternatives to current Tier 1 receptor binding, transactivation and uterotrophic assays (US EPA 2016). Regardless, the,

Table 3. Results from EDSP *in vitro* studies of 2,4-D (data from individual study reports, published in Coady et al. 2014).

Study	Assay and test system	Purity 2,4 D	Concentration range tested	Result	Klimisch score	Rationale for Klimisch score
LeBaron et al. 2011a	Estrogen receptor (ER) binding assay (US EPA 2009a) with uterine cytosol from Sprague Dawley rats	98.5%	10^{-11}–10^{-4} M	Negative, non-binder to ER	1	Guideline compliant study; dose range based on mammalian TK data to be below the TSRC
LeBaron & Kan 2011	Estrogen transcriptional activation assay (US EPA 2009b) with human ERα HeLa-9903 cells	98.5%	10^{-11}–10^{-4} M	Negative; no ER transactivation	1	Guideline compliant study; dose range based on mammalian TK data to be below the TSRC
LeBaron et al. 2011b	Androgen receptor (AR) binding assay (US EPA 2009c) with ventral prostate cytosol from Sprague-Dawley rats	98.5%	10^{-11}–10^{-4} M	Negative, non-binder to AR	1	Guideline compliant study; dose range based on mammalian TK data to be below the TSRC
Coady & Sosinski 2011	Aromatase assay (US EPA 2009d) with human recombinant aromatase and titrated androstenedione	98.5%	10^{-10}–10^{-4} M	Negative, non-inhibitor of aromatase activity	1	Guideline compliant study; dose range based on mammalian TK data to be below the TSRC
LeBaron et al. 2011c	Steroidogenesis assay (US EPA 2009e) with H295R cells	98.5%	10^{-10} 10 $^{-4}$ M	Small, significant 1 estradiol at 10^{-4} M; magnitude of change less than criterion used to define a positive response in validation studies (Hecker et al. 2008); considered negative	1	Guideline compliant study

proprietary methods used in ToxCast™ eliminate the opportunity to score study quality. However, ToxCast™ evaluates possible endocrine-receptor-related interactions in multiple assays, including evaluation of potential binding at both whole receptors and ligand-binding domains only, as well as examination of both agonist and antagonist activity in reporter-based systems (Judson et al. 2010). An evaluation of endocrine related ToxCast™ assays (Rotroff et al. 2013) demonstrated that:

> "ToxCast™ estrogen receptor and androgen receptor assays predicted the results of relevant EDSP Tier 1 assays with balanced accuracies of 0.91 ($p < 0.001$) and 0.92 ($p < 0.001$), respectively. Uterotrophic and Hershberger assay results were predicted with balanced accuracies of 0.89 ($p < 0.001$) and 1 ($p < 0.001$), respectively."

A more recent analysis (Cox et al. 2014) using a case study of HTS-derived models for predicting *in vivo* androgen, estrogen and thyroid endpoints showed that the more robust cross validation models (based on a set of endocrine ToxCast™ assays and guideline *in vivo* endocrine screening studies) have balanced accuracies from 79 to 85% for androgen or estrogen pathway interactions, but predicted substantially less accuracy for thyroid endpoints.

2,4-D purity was greater than 90% for all ToxCast™ assays. ToxCast™ assays for 2,4-D included:

Cell-free HTS assays. 2,4-D was tested at eight concentrations in the range of 0.00229–50 μM at receptor proteins of relevance to potential estrogen or androgen endocrine modulation. At the seven receptor proteins tested, inhibition of radio-ligand binding was less than 50% at all 2,4-D concentrations tested. This includes at the bovine and human ERs; bovine and human PRs; rat and human ARs; and human thyroid hormone receptor-α. Additionally, 2,4-D tested at eight concentrations in the range of 0.00914–20 μM exhibited less than 50% inhibition of human aromatase enzyme activity.

Cell-based HTS assays. In the cell-based HTS assays, 2,4-D was tested at 15 concentrations in the range of 0.0010–76.6 μM. Under these conditions, 2,4-D was considered inactive for agonist activity at the human AR, human ER-α and the human thyroid hormone receptor-ß. Furthermore, it did not block (or antagonize) the activity of established ligands at these receptor sites.

Multiplex transcriptian reporter assay. 2,4-D did not activate chimeric transcriptional proteins containing ligand-binding domains for the human AR, human ER-α, human estrogen-related receptor-α, human estrogen-related receptor-γ or human thyroid hormone receptor-α. The chemical also did not activate transcription at a human estrogen response element.

Aromatase. 2,4-D did not inhibit aromatase activity.

Thyroid. Although the thyroid pathway-related ToxCast™ assays for 2,4 D were negative, it should be noted that EPA has recently concluded that the ToxCast™ *in vitro* thyroid assays are not predictive of all relevant thyroid modes of

Table 4. Published *in vitro* assays relating to potential endocrine activity for 2,4-D.

Study	Assay and Test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Rationale for Klimisch score
Blair et al. 2000	Competitive rat ER binding assay in uterine tissue homogenates from ovariectomized Sprague-Dawley rats	99%	"2 high concentrations spanning 3 log concentrations"	Negative	3	Strengths: Well documented; method close to validated Guideline design; Weaknesses: 2,4-D specific data not provided; Only tested at 2 concentrations which were not reported
Fang et al. 2003	Binding to recombinant rat AR	Purity not specified; Supplier (Supelco) produces analytical standards but also mixtures	4.28×10^{-9}–4.28×10^{-4} M	Negative	3	Adequate study but lack of information on purity of 2,4-D
Jung et al. 2004	ER antagonist activity in yeast-based reporter system	"Highest grade commercially available"	Not specified	Negative	3	Weaknesses: concentrations tested not reported; possibly commercial formulation tested and purity unspecified. Not relevant/reliable: Yeast two-hybrid detection system
Jungbauer & Beck 2012	ER antagonist activity in yeast two-hybrid system	Not specified	Not specified	Negative	3	Weaknesses: concentrations tested not reported; purity unspecified. Not relevant/reliable: Yeast two-hybrid detection system
Kim et al. 2005	AR+ 22Rv1 cell proliferation with 2,4-D and its metabolite DCP	>98%	10^{-13}–10^{-5} M	Negative with 2,4 D or DCP alone, but positive with 2,4-D or DCP *with* dihydrotestos-terone (DHT) added	3	Weakness: Lack of solvent only control; data for DHT alone appears to have been generated for a single subset of tests only (set A in graphs); no rationale for 10 nM concentration of DHT added (possibly supra physiological); measured cell proliferation, which may be stimulated by ER-independent factors, including epidermal growth factor (based on information in the ATCC website (www.atcc.org) for this particular strain of cells. See also Sramkoski et al. 1999)
	Transactivation reporter assays with AR+ 22Rv1 and AR- PC3 cells	>98%	10^{-12}–10^{-6} M	Negative with 2,4-D or DCP alone, but positive with 2,4-D or DCP *with* DHT	3	Weakness: Lack of solvent only control; no rationale for concentration of DHT added (possibly supra physiological); only one concentration of 2,4-D tested in AR-PC3 cells.
	AR expression (mRNA and total protein levels) in 22Rv1 cells	>98%	10^{-7} M 2,4-D; 10^{-10} M DCP	Negative	3	Weakness: Single concentration of 2,4-D tested; Lack of solvent only control
	AR binding assay in monkey kidney COS cells	>98%	5×10^{-9} M to 5×10^{-7} M	Positive (50% inhibition with both 2,4-D and DCP)	3	Weakness: Lack of solvent only control; did not use reagent to separate unbound from bound ligand; evaluated limited number of test compound concentrations; typical dose-response

(*continued*)

Table 4. [illegible]ued

Study	Assay and Test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Rationale for Klimisch score
						curves for inhibition of receptor binding were not observed, even for the positive control, suggesting a potential solvent (vehicle) effect; biological plausibility of findings poor given lack of concordant findings in *in vivo* studies
	Nuclear translocation in PC3 cells	>98%	10^{-9} M	Negative with 2,4-D or DCP alone, but positive with 2,4-D or DCP *with* DHT	3	Weaknesses: Lack of solvent only control; only single concentration 2,4-D tested; no rationale for concentration of DHT added
Kojima et al. 2004	Estrogenic and anti-estrogenic activity in CHO cells transiently transfected with human ERβ	≥95%	10^{-8}–10^{-5} M	Negative	2	Strengths: Method very close to validated guideline; well documented; positive control used; weakness: specific response data for 2,4-D not provided.
	Androgenic and anti-androgenic activity in CHO cells transiently transfected with human AR	≥95%	10^{-8}–10^{-5} M	Negative	2	Strengths: Method very close to validated guideline; well documented; positive control used; weakness: specific response data for 2,4-D not provided
	Dual activity as ER agonists and AR antagonists	≥95%	10^{-8}–10^{-5} M	Negative	2	Well documented but method not formally validated. Weakness: specific response data for 2,4-D not provided
Lee et al. 2006	Estrogenic activity in yeast one-hybrid and two-hybrid systems	NR	10^{-7}–10^{-4} M	Negative in one-hybrid system; Positive in two hybrid system	3	Weakness: test material uncharacterized. Not relevant: Yeast one-hybrid and two-hybrid detection system
Lemaire et al. 2006	Estrogenic and anti-estrogenic activity in HeLa cells stably transfected with human ERα or human ER-β	>95%	10^{-6} M	Negative	3	Weakness: Only one concentration tested; otherwise adequate quality
Lin & Garry 2000	MCF-7 proliferation	2,4-D (reagent grade) 2,4-D isooctyl ester (reagent grade)	0.1–10 μg/mL	Negative	2	Weakness: MCF-7 cell line may provide variable responses and result may not be specific for estrogenicity (Odum et al. 1998)
	MCF-7 proliferation	2,4-D LV4 (commercial grade 66.24% 2,4 D isooctyl ester)); 2,4-D amine (commercial grade) 46.5% 2,4-D dimethylamine salt	0.1–10 μg/mL	Positive	3	Weaknesses: Results in commercial grade materials appear confounded due to formulation excipients as reagent grade materials did not show effects; MCF-7 proliferation may be a non-estrogen specific response and cell line may provide markedly variable responses (Odum et al. 1998)
Nishihara et al. 2000	Estrogenic activity using the yeast two-hybrid system	"Highest grade commercially available"	NR	Negative	3	Weaknesses: formulation and test material purity not characterized concentrations tested not reported; Not relevant: Yeast two-hybrid detection system
Orton et al. 2009	*Xenopus laevis* ovulation and ovarian steroidogenesis (production of progesterone,	>97%	6.25×10^{-6} and 62.5×10^{-6} M	Negative	3	Weaknesses: Methods not specific regarding stage of oocytes used (ranges given) or how

(*continued*)

Table 4. [illegible]ued

Study	Assay and Test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Rationale for Klimisch score
	testosterone and estradiol)					these were distributed to the test wells; methods not specific regarding number of replicates or criteria for accepting or rejecting replicate findings; non-validated assay; only two concentrations tested
	ER and AR agonist and antag onist activity in yeast	>97%	4.9×10^{-7}–1×10^{-3} M	Negative	3	Weaknesses: Not relevant: Yeast assay; otherwise well reported and conducted
Petit et al. 1997	VTG in mRNA expression in primary hepatocyte cultures derived from male rainbow trout	NR	10^{-4} M	Equivocal	3	Weaknesses. Only tested at a single concentration; test material uncharacterized VTG mRNA expression only 8% increase compared to control (cannot determine if ineffective or weakly responsive).
	β-galactosidase activity in yeast cells stably transfected with rainbow trout ER	NR	10^{-5}–10^{-4} M	Negative	3	Weakness: test material uncharacterized: Not relevant Yeast-based system
	ER competitive binding assay in yeast cells stably trans fected with rainbow trout ER	NR	$<10^{-4}$ M	Negative	3	Weaknesses: only tested at a single concentration; test material uncharacterized: Not relevant: Yeast-based system
Soto et al. 1995	Estrogenic activity by measuring MCF-7 proliferation	NR	NR	Negative	3	Weaknesses: test material uncharacterized and concentrations tested not reported; MCF-7 cell line may provide markedly variable responses (Odum et al 1998)
Sun et al. 2012	Estrogenic and anti-estrogenic activity in Vero cells	>99%	0.003–3.0 mg/L	Negative	2	Well-conducted and reported assay. Rationale for dose selection questionable; high dose exceeds potential human exposure although it falls within the linear TK range
	Androgenic and anti-androgenic activity in Vero cells	>99%	0.003–3.0 mg/L	Negative for androgenicity or anti androgenicity; at 3.0 mg/L increased the effects of testosterone (in anti-androgenic assay); other concentrations negative	2	Well-conducted and reported assay. Rationale for dose selection questionable, effect seen only at 3.0 mg/L which exceeds predicted concentrations for potential human exposure although it falls within the linear TK range; biological relevance of the finding questionable as the test was designed to measure anti-androgenic activity rather than potentiation or androgenic activity
	Agonist and antagonist activity to TR in Vero cells	>99%	0.003–3.0 mg/L	Negative	3	Well-conducted and reported assay; however *In vitro* thyroid assay model not validated
Vonier et al. 1996	Competitive binding to ER and PR extracted from the oviduct tissues of adult female alligators	>99%	NR	Negative	3	Well-conducted study with positive control; Weaknesses: concentrations tested not reported.

action in *in vivo* studies (Rotroff et al. 2013), concordant with findings in Cox et al. 2014.

Reif et al. (2010) provides a "Tox-pi" diagram for 2,4-D which confirms that the full range of endocrine-related ToxCast™ assays for 2,4-D were negative.

Based on the multiple assays and consistency of results with *in vivo* and *in vitro* studies of 2,4-D, the ToxCast™ program endocrine-relevant results are considered supportive for concluding 2,4-D is not likely to show potential interactions with either the estrogen or androgen pathways. These data are consistent with the regulatory (EDSP Tier I) *in vitro* data and with the majority of *in vitro* studies of 2,4-D in the published literature.

Ecotoxicological studies of 2,4-D relevant to assessment of potential endocrine pathway interactions

The AMA and FSTRA conducted to meet EDSP Tier 1 screening requirements, and a one-generation reproductive toxicity study in quail (Mitchell et al. 2000) conducted to meet prior regulatory testing requirements, provide the most relevant and substantive ecotoxicological data for studying the possible endocrine activity of 2,4-D. These studies are briefly summarized below for ready reference. Specific data are provided for the quail study because this study was unpublished and it is the only relevant bird study identified. Further details on these studies may be found in Supplementary Appendix III and, for the frog and fish assays, in Coady et al. 2013. Regulatory and published studies with Klimisch scores of 2 or higher that are relevant to evaluation of potential endocrine interactions are summarized in Supplementary Appendix III. Studies with lower Klimisch scores, or those adequate studies that were found not to include relevant endpoints, are summarized in Supplementary Appendix VI.

Coady et al. 2010 (published in Coady et al. 2013)

The study design of the AMA (Coady et al. 2010) corresponded with guidelines: OPPTS 890.1100 (US EPA 2009l) and OECD 231. In brief, African clawed frog (*Xenopus laevis*) tadpoles were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days. Nominal test concentrations of 0, 0.4, 4, 40 and 100 mg/L were tested, with the high concentration selected based on prior acute toxicity studies, and equivalent to a limit concentration in the guideline. Concentrations were monitored over the course of the study. Although decreases from the nominal concentration (probably due to biodegradation) were noted, particularly at the lowest concentration tested, the concentrations tested were well documented and the study is considered valid with a Klimisch score of 1 for this guideline compliant study. The mean measured concentrations of 2,4-D in this assay were 0.273, 3.24, 38.0 and 113 mg/L for the 0.4, 4, 40 and 100 mg/L nominal concentrations, respectively.

There was no indication of systemic toxicity in this study, with no effects on survival, clinical signs or body weights (evaluated days 7 and 21). There were no effects on days 7 or 21 on snout-vent length, Nieuwkoop and Faber (1994) developmental stage, hind limb length (normalized to snout-vent length or asynchronous development). There were no exposure-related findings on histopathological evaluation of the thyroid following necropsy on day 21. In sum, there were no biologically significant exposure-related effects on the thyroid gland or the morphological endpoints of this assay under thyroid control (hind limb length and developmental stage), and there is no evidence of a potential interaction with the HPT axis in this Tier 1 EDSP AMA tested to the assay limit concentration of 100 mg 2,4-D/L.

Marino et al. 2010 (published in Coady et al. 2013)

Marino et al. 2010 tested 2,4-D in a FSTRA (US EPA 2009g). The study was conducted in compliance with OPPTS 890.1350 and OECD 229. Sexually mature fathead minnows (*Pimepholes promelas*) were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days at nominal concentrations of 0, 0.4, 4, 40 and 100 mg/L. The high concentration was selected based on acute toxicity tests and an early life stage test with fathead minnows (Alexander et al. 1983; Mayes et al. 1990); and also represents a limit concentration for the assay. The negative control was untreated laboratory dilution water. Although decreases from the nominal concentration (probably due to biodegradation) were noted, particularly at the two lowest concentrations tested, the concentrations tested were well documented and the study is considered valid with a Klimisch score of 1 for this guideline-compliant study.

Results are summarized in Table 5. The only statistically significant finding compared to the controls was a decrease in fecundity (considered a non-specific finding) among fish exposed to the highest concentration of 2,4-D. In the absence of effects upon other more specific endocrine-mediated endpoints, the isolated effect on fecundity at 100 mg a.i./L is considered most likely to reflect systemic toxicity and a generalized stress response. This concentration is relatively high (approximately 1/3 of the acute LC50 value in fish), is the limit concentration for the FSTRA, and is a concentration which exceeds the maximum acceptable toxicant concentration (MATC) for larval fish survival in an early life stage toxicity test with fathead minnows (Mayes et al. 1990).

In conclusion, 2,4-D does not appear to interact with the estrogen, androgen or steroidogenic pathways, or with the HPG axis in fathead minnows tested up to the limit concentration in this EDSP Tier 1 FSTRA.

Mitchell et al. 2000

An avian single generation reproductive toxicity study (Mitchell et al. 2000) of 2,4-D showed no systemic toxicity to quail and a lack of potential endocrine-related effects. This study complied with Federal Insecticide, Fungicide and Rodenticide Act (FIFRA) Guideline 71–4 and OECD Guideline 206 and was conducted under GLP, and is therefore scored a Klimisch 1.

Table 5. Summary: fish short-term reproduction assay with 2,4-D (Marino et al. 2010).

Nominal concentration of 2,4-D (mg a.e./L)	0.4	4.0	40	100	
Mean measured concentration of 2,4-D (mg a.e./L)	0.245	3.14	34.0	96.5	Interpretation
Systemic toxicity					
Survival	N	N	N	N	No effect. One high dose death, not likely to be exposure-related.
Clinical signs	N	N	N	N	No effect. No abnormal behavior or coloration was observed.
Body weight M	N	N	N	N	No effect
Body weight F	N	N	N	N	No effect.
Feed consumption	N	N	N	N	No observations noted.
Body length M	N	N	N	N	No effect.
Body length F	N	N	N	N	No effect
Fertilization success	N	N	N	N	No effect; no difference between groups in numbers of fertilized eggs.
Fecundity	N	N	N	↓	May be associated with stress; high concentration exceeds larval maximum acceptable toxicant concentration (MATC) in an early life stage study with *P. promelas*
Estrogen pathway – potentially indicative endpoints					
Nuptial tubercles M	N	N	N	N	No effect in M; present in males in equivalent numbers between control and dosed groups.
Gonadal somatic index (GSI) M	N	N	N	N	No effect.
GSI F	N	N	N	N	No effect.
VTG M	N	N	N	N	No increase VTG in males.
VTG F	N	N	N	N	No increase or decrease VTG in females.
Gonadal histopathology	N	N	N	N	No effect in M or F. The pattern of predominant germ cell distribution (staging) in testes and ovaries was comparable between the controls and all dosed groups.
Androgen pathway – potentially indicative endpoints					
Nuptial tubercles M	N	N	N	N	No effect in M; present in males in equivalent numbers between control and dosed groups
Nuptial tubercles F	N	N	N	N	Not present in females normally; none observed.
GSI M	N	N	N	N	No effect.
GSI F	N	N	N	N	No effect.
VTG M	N	N	N	N	No alteration in M VTG.
VTG F	N	N	N	N	No alteration in F VTG.
Gonadal histopathology	N	N	N	N	No effect in M or F. The pattern of predominant germ cell distribution (staging) in ovaries and testes was comparable between the controls and all dosed groups

N: no effect; M: male; F: female.
↓ Statistically significant decrease ($p < 0.05$) compared to control.

2,4-D acid (96.9% pure) was administered to adult Northern Bobwhite male and female quail (*Colinus virginianus*) for 21 weeks via the diet at 0, 160, 400 and 1000 ppm. The high dose level complies with the limit dose recommended in OECD Guideline 206. The no-observed effect concentration for northern bobwhite quail exposed to 2,4-D acid in the diet during the study was 1000 ppm, the highest concentration tested. There were no effects on mortality, clinical signs, body weight or feed consumption of adult birds and no exposure-related findings at necropsy. Two high-dose deaths were attributable to injury. Other results are summarized in Table 6 below. Slight but statistically significant decreases in the percent of hatchlings/eggs set and 14/day survivors/eggs set and a non-statistically significant decrease in the mean percent of viable embryos as a percent of eggs set were observed at the low dose. These findings were attributable primarily to one pen, in which no eggs were fertile, and the male showed quiescent testes at necropsy. This fact and the lack of dose response led to the conclusion that this finding was not exposure-related. Eggshell thickness was statistically significantly increased at 400 ppm; primarily attributable to results from one pen with an elevated eggshell thickness. This finding was not considered exposure related based on the slight nature of the finding, the attribution to one pen and the lack of dose response. This study is considered valid; it predicts a very low hazard of reproductive toxicity of 2,4-D to birds and a low likelihood of endocrine-related effects on birds.

Review of studies for study quality

The EDSP Tier 1 ecotoxicological studies, one-generation quail and published ecotoxicological studies identified as possibly relevant to assessment of potential endocrine pathway interactions are tabulated in Table 7. There were no findings in the regulatory toxicological studies considered likely to reflect endocrine pathway interactions. A series of studies by Crain et al. (1997; 1999) was considered valid; these studies using 2,4-D applied to alligator eggs was validated with a positive control and showed no effects of 2,4-D (summarized in Supplementary Appendix III). Other studies are summarized in Supplementary Appendix VI.

Table 6. Results in one-generation quail reproductive toxicity (Mitchell et al. 2000).

Dose (ppm)	Control	160	400	1000
Number of replicates	16	16	16	14
Total eggs laid	746	788	768	638
Eggs laid/maximum laid (%)	75	80	77	74
Viable embryos/eggs set (%)	91	76	95	91
Live 3-week embryos/viable embryos (%)	99	98	99	99
Hatchlings/live 3-week embryos (%)	95	91	92	94
14-Day survivors/hatchlings (%)	94	96	95	93
Hatchlings/eggs set (%)	85	68*	86	86
14 Day survivors/eggs set (%)	80	65*	82	80
Hatchlings/maximum set (%)	64	55	67	60
14-Day survivors/maximum set (%)	60	52	64	56
Mean (± SD) eggshell thickness (mm)	0.226 ± 0.016	0.228 ± 0.010	0.239* ± 0.015	0.232 ± 0.232
Eggs cracked/eggs laid (%)	1	2	1	3
Mean (± SD) body weight hatchlings (g)	6 ± 0	6 ± 1	6 ± 0	6 ± 0
Mean (± SD) body weight 14-day survivors (g)	26 ± 2	26 ± 3	26 ± 2	26 ± 2

*$p < 0.05$

Mammalian toxicological studies of 2,4-D relevant to assessment of potential endocrine pathway interactions

Selection of regulatory mammalian toxicology studies for inclusion in review

Regulatory (unpublished) mammalian toxicity studies of 2,4-D acid conducted for pesticide registration purposes were reviewed. Fourteen studies with the most relevant endpoints for evaluation of potential endocrine toxicity and most comprehensive reporting were selected for this WoE evaluation. Data from several of these studies have also been published; citations to both the reports and publications are provided.

These studies included, most critically, an EOGRT study of 2,4-D (Marty et al. 2010 published in Marty et al. 2013), which, as noted, used TK data to inform dose selection, and which included multiple endpoints specifically added in consultation with the US EPA and the Canadian PMRA to provide additional information on potential endocrine interactions of 2,4-D with the estrogen, androgen or thyroid pathways. At the time this study was conducted, the guideline for an EOGRT study was still under development; however, based on the extensive vetting of the study design, similarity to the adopted test guideline, and involvement of two regulatory authorities in both the study design and critical decision points, it met all the objectives of the current OECD (2012a) study guideline (443).

Other selected regulatory studies include:

- US EPA Office of Pesticide Program (OPP) 83-4 guideline two-generation reproductive toxicity study (Rodwell & Brown 1985);
- OPP 83-3 guideline developmental toxicity study in rats (Rodwell 1983; Charles et al. 2001);
- OPP 83-3 guideline developmental toxicity study in rabbits (Hoberman 1990; Charles et al. 2001);
- OPP 82-1 guideline 13-week rat subchronic toxicity study (Schulze 1991a; Charles et al. 1996a)
- non-guideline 13-week rat subchronic toxicity studies (Gorzinski et al. 1981a, 1981b);
- OPP 83-5 guideline two-year rat chronic toxicity/oncogenicity study (Jeffries et al. 1995; Charles et al. 1996c);
- OPP 82-1 guideline 13-week mouse subchronic toxicity (Schulze 1991b);
- OPP 83-2 guideline-equivalent mouse oncogenicity evaluations (Stott 1995a, 1995b; Charles et al. 1996c);
- OPP 82-1 non-guideline 13-week dog subchronic toxicity study (Schulze 1990)
- OPP 82-1 guideline 13-week dog subchronic toxicity (Dalgard 1993a; Charles et al. 1996b); and
- OPP 83-1 guideline dog chronic toxicity study (Dalgard 1993b; Charles et al. 1996b).

The regulatory mammalian toxicological database provides an opportunity to evaluate consistency of responses across species and strains, and also across exposure durations.

The Tier 2 EDSP-equivalent EOGRT study (Marty et al. 2010; published in Marty et al. 2013) and the two-generation reproductive study findings (Rodwell & Brown 1985) are briefly summarized in this section because these two studies provide by far the most comprehensive and relevant endpoints for evaluating the potential EAT and steroidogenesis interactions of 2,4-D.

Subsequently, other subchronic and chronic mammalian regulatory studies and studies identified in the published literature are tabulated and scored for study quality. Further details on these studies may be found in Appendix IV for studies considered to meet Klimisch criteria 1 or 2, and in Appendix VII for studies considered to meet Klimisch criteria 3 or 4.

The primary caveat regarding regulatory studies other than the EOGRT, and the majority of the published mammalian toxicological studies is that the high dose level was based on (or in some cases exceeded) a classic maximum tolerated dose (MTD) and far exceeds the TSRC. High dose level findings above the TSRC are presented but, as discussed in the Introduction, are not considered relevant for human hazard characterization or risk assessment.

Additionally, as discussed previously, information from the dog studies is not considered relevant for human risk assessment because the dog lacks an effective organic acid renal transport mechanism (Timchalk 2004); however, data from the dog studies are included because they may be useful in predicting potential effects on other species lacking an

Table 7. 2,4-D ecotoxicological studies possibly relevant to assessment of potential endocrine interactions.

Reference	Assay and test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Klimisch score rationale
Amphibians						
Coady et al. 2010; (Coady et al. 2013)	Amphibian metamorphosis assay (US EPA 2009t, OECD 231). African clawed frog tadpoles exposed in continuous flow-through system for 21 days.	98.6%	0.4–100 mg a.e./L	No exposure-related effects	1	Guideline and GLP compliant study; tested to limit concentration
Aronzon et al. 2011	South American toad exposed to 2,4-D DBE or formulated product either through embryogenesis or in pulsed exposures	99% 2,4-D di-butyl ether	1–15 mg/L 2,4-D DBE for continuous exposure	"Teratogenic" to toads	3	No controls were included in this study; results cannot be interpreted
Heggstrom 2009	Wood frog tadpoles in microcosms exposed to 2,4-D dimethylamine	99% 2,4-D dimethylamine	0.1–100 µg/L	Negative for survival, deformities, effects on time to metamorphic climax; total length decreased in mid dose group only; no effect on plasma corticosterone levels	3	High non-exposure related mortality
Heggstrom 2009	Wood frog tadpoles in field ponds (agricultural and forested) exposed to 2,4-D dimethylamine	99% 2,4-D dimethylamine	10 µg/L	Tadpoles from the agricultural pond in the absence of 2,4-D dimethylamine and tadpoles from the agricultural pond applied with 10 µg/L 2,4-D dimethylamine gave similar results in the measured endpoints	3	Control tadpoles from the two ponds showed some marked differences; only one concentration level evaluated
Stebbins-Boaz et al. 2004	*Xenopus* oocytes exposed to 2,4-D sodium salt	NR	0.6–2.43 g/L (2.5–10 mM)	Germinal vesicle breakdown inhibited	3	Mechanistic study; extremely high doses; purity not reported
LaChapelle et al. 2007	*Xenopus* oocytes exposed to 2,4-D	NR	2.43 g/L (10 mM)	Irreversible dysfunction of meiotic signaling	3	Mechanistic study; only one extremely high dose evaluated; purity not reported
Morgan et al. 1980	Frog embryo teratogenic assay in *Xenopus*	Commercial formulation (99%)	180–270 mg/L	Teratogenic only at high concentrations	2	Adequate study but irrelevant to evaluation of potential endocrine effects; formulation tested
Vardia et al. 1984	Indian tadpoles exposed to 2,4-D in a static renewal exposure system	NR	7.5–11 mg/L	96-h LC_{50}: 8.05 mg/L	4	Experiment not described in detail; lacks information on potential endocrine effects
Fish						
Marino et al. 2010; (Coady et al. 2013)	Fish short term reproduction assay (OPPTS 890.1350, OECD 229, US EPA 2009g). Adult fat head minnows were exposed via continuous flow-through for 21	98.6%	0.4–100 mg a.e./L	No interaction with endocrine pathways; decreased fecundity at highest concentration tested attributed to systemic toxicity	1	Guideline and GLP compliant study; tested to limit concentration

(continued)

Table 7. Continued

Reference	Assay and test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Klimisch score rationale
	days.					
Holcombe et al. 1995	Acute larval survival and growth tests in Japanese Medaka exposed to 2,4-D acid	99% 2,4-D acid	567–8970 mg/L	96-h LC_{50}: 2780 mg/L 2,4-D Acid	2	Adequate study but no specific endocrine endpoints evaluated
Holcombe et al. 1995	Chronic larval survival and growth tests in Japanese Medaka exposed to 2,4-D acid	99% 2,4-D acid	27.2–425 mg/L	Survival and growth reduced at 56.5 mg/L	2	Adequate study but no specific endocrine endpoints evaluated
Holcombe et al. 1995	Chronic larval survival and growth tests in Japanese Medaka exposed to 2,4-D acid	99% 2,4-D acid	2.37–60.2 mg/l	Survival and growth reduced at 60.2 mg/L	2	Adequate study but no specific endocrine endpoints evaluated
Koç & Akbulut 2012	Acute toxicity to ovary of zebrafish	NR	0.1 1 mg/L	Ovarian histopathological changes and atretic follicles following 5 days exposure	3	Source and purity not defined; report says 2,4-D "available as a formulation" so possible a formulation was tested; methods not well defined; potential sectioning artifacts in slides
Padilla et al. 2011	Zebrafish developmental screening assay in zebrafish embryos exposed to 2,4-D	>90%	0.001–80 μM	Negative in concentration range study; weak positive in single high concentration study	2	Reported findings too non-specific to be useful for WoE
Rehwold et al. 1977	Acute toxicity tests with striped bass, the banded killifish, pumpkinseed, white perch, American eel, carp, and guppy exposed to 2,4-D	NR	NR	96-h LC_{50} ranged from 26.7 mg/L (banded killifish) to 300.6 mg/L (American eel)	4	Experimental detail not provided; fish field collected
Rehwold et al. 1977	Chronic toxicity tests with striped bass, the banded killifish, pumpkinseed, white perch, American eel, carp, and guppy exposed to 2,4-D	NR	0.1 mg/L	No observable physiological symptoms	4	Experimental detail not provided; fish field collected
Rehwold et al. 1977	Breeding effects on guppy chronically exposed to 2,4-D	NR	0.1 mg/L	No observable physiological symptoms	4	Experimental detail not provided
Xie et al. 2005	Juvenile rainbow trout exposed to 2,4-D dimethylamine in static test system	NR	0.00164–1.64 mg/L	VTG levels significantly greater at 0.164 and 1.64 mg/L	3	High variability; small sample size, sex of fish not determined, purity of test material not reported
Avian						
Mitchell et al. 2000	Avian single generation reproductive study in Northern Bobwhite quail. US EPA OPP 71-4;	96.9%	160, 400, 1000 ppm (No mg/kg/d dose calculated)	No treatment-related effects including on potentially endocrine related parameters	1	Guideline compliant

(*continued*)

Table 7. Continued

Reference	Assay and test system	Purity 2,4-D	Concentration range tested	Result	Klimisch score	Klimisch score rationale
	OECD Guideline 206; Subjects exposed via diet for 21 weeks.					
Somers et al. 1974	Fertilized hen's eggs exposed to 2,4-D by spraying	NR	2.8–3.4 kg/ha and 44.8 kg/ha	No mortalities *in ovo*, at pip, or at hatch; no weight gain effects	3	Non-conventional method of application; purity not reported; no positive control; evidence of poor absorption into the egg content
Somers et al. 1978a	Hens and cockerels exposed to 2,4-D by spraying	Formulation purity NR	111.2 kg/ha	No effect on chicken reproduction in various endpoints measured	3	Non-conventional method of application; formulation; single concentration and purity not reported; no positive control; evidence of poor absorption into the egg content
Somers et al. 1978b	Hens eggs exposed to 2,4-D by spraying	Formulation purity NR	111.2 kg/ha	No effect on hatching success, weight gain, or mortality;	3	Non-conventional method of application; formulation; single concentration; purity not reported; no positive control; evidence of poor absorption into the egg content
Reptiles						
Crain et al. 1997	American alligator eggs exposed topically to 2,4-D prior to sexual differentiation	97.6%	0.14–14 ppm	No effects on sex reversal, plasma steroid concentrations, gonadal aromatase activity	2	Non-conventional study design; validated with estradiol positive control
Crain et al. 1999	American alligator eggs exposed topically to 2,4-D prior to sexual differentiation	97.6%	0.14–14 ppm	No effects on hepatic aromatase activity or testicular histopathology	2	Non-conventional study design; validated with estradiol positive control
Spiteri et al. 1999	American alligator eggs exposed topically to 2,4-D prior to sexual differentiation, at two temperatures	97.6%	0.14–14 ppm	No effects on sex reversal, hepatic aromatase activity or gonadal histopathology	2	Non-conventional study design; validated with estradiol positive control
Mixed species						
Relyea 2005	Algal microcosms and other microcosms containing 25 species exposed to 2,4-D	Formulation (44.5%)	0.117 mL/m^2	No effect on community diversity, survival, or biomass	3	Formulation tested and single concentration; No specific endocrine endpoints evaluated

NR: not reported.

effective OAT 1 transporter (if any) in the environment and because they provide information suggesting that the thyroid findings for 2,4 D may be rodent-specific.

Marty et al. 2010 (published in Marty et al. 2013)

The EOGRT study of 2,4-D (Marty et al. 2010, published in Marty et al. 2013) was specifically designed to provide sufficient information to assess whether endocrine targets are, in fact, altered with *in vivo* exposure, and to provide the basis for robust risk assessment of 2,4-D, including risk assessment protective for any potential endocrine effects. This study design provides a reliable basis for establishing the potential of 2,4-D to interact with the estrogen, androgen or thyroid pathways and is considered a Tier 2-equivalent EDSP assessment. The OECD (2012b) considers the EOGRT study a preferable method for evaluation of *in vivo* endocrine disruption in that it evaluates endocrine-sensitive endpoints not found in conventional 2-generation bioassays. This study is assigned a Klimisch score of 1.

As discussed in the Introduction, this study used extensive TK information on 2,4-D to set doses. Based on blood levels obtained during the EOGRT study, the high dose in males (800 ppm) adequately approximated or slightly exceeded the TSRC, but the high dose in females (600 ppm) clearly exceeded the TSRC.

A summary of the study design and discussion of results and endocrine-related parameters is included in Appendix IV.

EOGRT study key parameters and findings are summarized in Table 8.

In conclusion, there was no evidence of adversely altered endocrine function in a comprehensive EOGRT study of 2,4-D. Slight adaptive effects were seen on thyroid hormone homeostasis at the high dose in a single life-stage, at a dose exceeding the TSRC and not relevant for human risk assessment.

Rodwell and Brawn, 1985

This study was a two-generation OPP 83-4 Guideline reproductive toxicity study in Fischer 344 rats. 2,4 D (97.5% purity) was administered in the diet at nominal dose levels of 0, 5, 20 and 80 mg/kg/day (30/sex/dose) for one full generation and at 0, 5 and 20 mg/kg/day for the second generation. The 80 mg/kg/day group was dropped after the first generation because it exceeded a MTD, based on excessive mortality among the F1b pups following a mis-dosing during gestation and lactation. The mis-dosing resulted in all groups of F1b dams and pups being exposed to greater than nominal doses; high-dose dam exposure was >100 mg/kg/day. There was no dose concentration adjustment in this study and the high dose exceeded the TSRC. Because of the mis-dosing and several study deficiencies, this study is scored a Klimisch score of 2. Details on the study are provided in Supplementary Appendix IV; a summary of key parameters evaluated and results for the Rodwell and Brown (1985) study are presented in Table 9.

In summary, there were no robust indications of interaction with the estrogen or androgen pathways in this study; thyroid function was not evaluated.

A summary of regulatory developmental, subchronic and chronic toxicity studies follows in Table 10, and published mammalian toxicological studies in Table 11.

The key points from the mammalian regulatory developmental, subchronic and chronic toxicity studies (Table 10) and published toxicological studies (Table 11) are:

- 2,4-D has not been shown to have exposure-related changes in endpoints potentially related to EAT pathway or steroidogenesis endpoints at doses below the TSRC in high quality studies, including in a comprehensive EOGRT. Findings at higher (in most cases much higher) doses are not considered relevant to human risk assessment.
- No data provide robust indications of interactions with the estrogen or androgen pathways or with steroidogenesis. Adaptive effects on thyroid parameters are seen at doses exceeding the TSRC in rodents, but not in dogs.
- Developmental and subchronic toxicity studies of 2,4 D esters and amines show no unique endocrine related toxicity and results are generally consistent and predictable based on the acid studies. Note that these compounds break down rapidly to the acid form.

Occupational and epidemiological investigations

Male reproductive health

Lerda and Rizzi, 1991

Thirty-two farmers occupationally exposed to 2,4 D and 25 non-exposed controls were studied for the following reproduction-related effects: ejaculatory volume, sperm count, sperm motility and sperm morphology. Exposure level was estimated by measuring the concentration of 2,4-D in the urine. Mean 2,4-D concentrations were 9.02 milligrams per liter (mg/L) in the exposed group, while 2,4-D was not detectable in the control group.

The investigators reported that the incidence of astheno spermia, necrospermia and teratospermia were greater in the exposed group, and that sperm motility was decreased. However, many study-specific details were not reported, including background of controls, number of participants excluded due to spermatogenesis-affecting health conditions, method used in "consideration" of external factors, detection limit for 2,4-D in urine, time period of urine collection and ranges of sperm parameters and 2,4-D urine levels evaluated (only means provided in the paper). The selection of controls appeared inappropriate, as comparison was to workers in the field exposed to 2,4-D, but the controls were not agricultural workers or doing similar field work. Field work involves exposure to other factors (e.g. increased temperature, dusts and allergens) that could potentially alter sperm parameters. In addition, no attempt was made to correlate 2,4-D urine levels with any specific sperm parameter changes or anomalies. Therefore, this study is considered too limited in scope and relevant details, and is not considered to provide reliable

Table 8. F1-extended one generation dietary toxicity study summary table (Marty et al. 2010).

Dose (ppm) Targeted mg/kg/day doses	100 5	300 15	600F/800 30F/40M	Interpretation
Systemic toxicity				
Mortality	N	N	N	No exposure-related mortality in any group
Live birth index	N	N	N	No effect
Mean number of pups	N	N	N	No effect
Clinical signs	N	N	N	No effect
Body weight	N	N	↓	Decreased P1 F bw during lactation; decreased F1 pup weight
Feed and water consumption	N	N	↓P1 F	Decreased P1 F feed consumption during lactation
Liver weight	N	N	N	No effect
Kidney weight	N	↑F1 F	↑P1 M and F; F1 F	Exposure related; very slight and considered non-adverse at 300 ppm
Kidney histopathology	N	F1 M	P M; F1 M and F	Exposure related; very slight and considered non-adverse at 300 ppm
Estrogen pathway-potentially mediated endpoints				
Sexual maturation (vaginal opening)	N	N	N	No effect F1
Ano-genital distance	N	N	N	No effect F1
Nipple retention (male)	N	N	N	No effect F1
Estrous cycling	N	N	N	No effect P or F1
Female mating	N	N	N	No effect P
Mean duration of gestation	N	N	N	No effect P
Gestation index	N	N	N	No effect P
Pup sex ratio	N	N	N	No effect F1
Ovaries (paired) weights	N	N	N	No effect P or F1
Differential ovarian follicle count	N	N	N	No effect F1
Uterus weight w oviducts and cervix	N	N	↑(N)	Non-statistically significant; increase within HCD; stage of estrus uncontrolled at necropsy; estrous cyclicity not changed; conclusion no effect P or F1
Uterine histopathology	N	N	N	No effect P or F1
Ovarian histopathology	N	N	N	No effect P or F1
Vaginal histopathology	N	N	N	No effect P or F1
Androgen pathway-potentially mediated endpoints				
Sexual maturation (preputial separation)	N	N	↓(N)	Slight delay F1 M attributed to decreased body weight before and post weaning (artifact from group assignments)
Sperm parameters	N	N	N	No effect P or F1
Male fertility	N	N	N	No effect P
Epididymis weight	N	N	N	No effect P or F1
Testicular weight	N	N	N P ↓(N) P F1 pups; N F1 adults	No effect P or F1 adults Decreased testis weights in F1 PND 21 pups at high dose strongly correlated with decreased body weight; did not persist in adults
Prostate weight	N	N	↓(N) P1; N F1	P1 prostate weights not statistically different from control; decreases not considered exposure related because the absolute and relative prostate weights in the control group were atypical, exceeding the laboratory HCD ranges; not seen in F1
Seminal vesicles with coagulating glands weight	N	N	↓(N) P1; N F1	P1 males decreased seminal vesicle weight not considered exposure-related because the absolute and relative seminal vesicle weight in the control group were atypical, exceeding the laboratory HCD ranges; not seen in F1
Testicular histopathology	N	N	N	No effect P or F1
Prostate histopathology	N	N	N	No effect P or F1
Epididymides histopathology	N	N	N	No effect P or F1
Seminal vesicle histopathology	N	N	N	No effect P or F1
Coagulation gland histopathology	N	N	N	No effect P or F1
Thyroid pathway-potentially mediated endpoints				
T3	N	N	↓GD17 satellite F N Other life stages	No adverse findings; response in dams considered adaptive
T4	N	N	↓GD17 satellite F N Other life stages	No adverse findings; response in dams considered adaptive
TSH	N	N	↑GD17 satellite F N Other life stages	No adverse findings; response in dams considered adaptive
Thyroid weight	N	N	N	No adverse findings

(continued)

Table 8. Continued

Dose (ppm) Targeted mg/kg/day doses	100 5	300 15	600F/800 30F/40M	Interpretation
Thyroid histopathology	N	N	Colloid depletion GD 17 F N Other life stages	Adaptive response in GD 17 satellite female dams consistent with decreased thyroxine at high dose; not considered adverse
DNT effects (Functional observational battery, motor activity, startle, histopathology)	N	N	N	No effects (assessed in F1 cohort)
Brain morphometry or myelination	N	N	N	No effects (assessed in F1 cohort)
Other potentially endocrine-mediated endpoints				
Pituitary weights	N	N	↓(N) F1 Set 3 M	Not considered exposure related; no findings in any other set or correlating effects, including the absence of histopathological changes
Pituitary histopathology	N	N	N	No effect P or F1
Adrenal (paired) weights	N	N	N	No effect P or F1
Adrenal histopathology	N	N	N	No effect P or F1

NA: not applicable; N: no effect; (N): finding not considered related to potential endocrine pathway interaction, see interpretation; M:male; F: female; P1: parental generation, F1: first generation (P offspring); GD: gestation day.
*Increase relative to control
↓Decrease relative to control

evidence of male reproductive toxicity or endocrine disruption resulting from occupational exposure to 2,4 D.

Garry et al. 2001

Twenty-four applicators and 15 minimally exposed foresters (control subjects) were studied for biomarker outcomes compared to urinary levels of 2,4-D. Categorized by applicator method, men who used hand-held, backpack sprayer applicators showed the highest average level (453.6 ppb) of 2,4-D in urine. No significant differences in follicle-stimulating hormone (FSH), total testosterone or free testosterone levels between application methods were reported. Significantly increased luteinizing hormones (LH) levels were reported in backpack applicators and boom-sprayer applicators combined; however, no significant effect on LH levels was observed in either backpack applicators or boom-sprayer applicators alone.

No correlation was shown between FSH, free testosterone or total testosterone concentrations with 2,4-D urinary levels at the time of maximum 2,4 D usage. In contrast, LH levels were reported to correlate with 2,4-D urinary levels at the time of maximum 2,4-D usage (using 21 of 24 applicators). LH levels are subject to considerable inherent variation and single samples from individuals are unlikely to provide a reliable profile (Partsch et al. 1994). Total testosterone levels after the application season were reported to correlate with 2,4-D urinary levels at the time of peak 2,4-D use. The study authors acknowledged that the limited sample size warrants cautious interpretation of the data. This study is considered too limited in scope to provide substantive evidence of endocrine modulation caused by exposure to 2,4-D.

Swan et al. 2003

Swan et al. (2003) in a case-control study evaluated semen quality, sperm concentration, morphology and motility in general population participants in two states, evaluating levels of pesticide metabolites taken close to the time of sample collection as surrogates for exposure. They found no statistically significant effects on sperm concentration or quality, or increased abnormal sperm at urinary 2,4-D levels above the limit of detection (LOD). (It should be noted that very few samples were above the LOD for 2,4-D.) The authors commented that the results for 2,4-D should be "considered borderline, with small and somewhat inconsistent associations." The abstract to the paper indicates that 2,4-D was "associated with poor semen quality in some analyses." Based on the results presented in the paper this statement in the abstract appears speculative, and unsupported by the data, unless the "analyses" in the statement refers to one for all pesticides combined, and not 2,4 D specifically. This study does not provide any robust evidence that 2,4-D exposure is associated with poor semen quality in humans. It is considered too limited, due to the low numbers of control and case subjects with urinary 2,4-D levels above the LOD, to be considered in the WoE as evidence for presence or absence of an association.

Thyroid

Knopp, 1994

The urinary excretion of 2,4-D was measured during eight biological monitoring studies over a five-year period (1985–1989) of 27 men and 18 women employees exposed during the production and formulation of 2,4-D and related sodium and dimethylamine salts (DMAs). In addition, venous blood samples were collected in three legs of the studies, and thyroid hormone concentrations in blood were measured.

Results showed that 2,4-D was detectable in serum and urine of all persons, but in varying amounts. The highest urinary concentration was 19.5 ppm, and the 2,4-D urinary concentration profile for a weekly interval showed an increase in exposure during the work week.

No notable abnormalities of thyroid hormone concentrations in blood were found. It should be noted, however, that

Table 9. 2,4-D: two-generation reproductive toxicity study summary table (Rodwell & Brown 1985).

Dose (mg/kg/day)	Vehicle control	5	20	80*	Interpretation
Systemic toxicity					
Mortality	N	N	N	Y	Pup mortality increased in the F1b litters (overdosed during mating, gestation and lactation).
Mean number of pups	NA	N	N	N	No effect
Live birth index	NA	N	N	Y	Increased stillbirths in F1b litters (overdosed during mating, gestation and lactation).
Clinical signs	NA	N	N	N	No effect
Body weight	NA	N	N	↓	Decreased body weights in F1a and F1b litters at ~80 mg/kg/day
Feed consumption	NA	N	N	↓	Decreased in dams producing F1b litters (overdosed during mating, gestation and lactation).
Liver weight	NA	N	N	N	No dose related effect
Kidney weight	NA	N	N	N	No dose-related effect
Estrogen pathway –potentially indicative endpoints					
Estrous cycling (extrapolated)	NA	N	N	N	No effect (extrapolated from time to mating)
Mating index	NA	N	N	N	No effect
Fertility index	NA	N	N	↓	Decreased (not statistically significantly) in mating to produce F1b litters at ~80 mg/kg/day (overdosed during mating, gestation and lactation).
Mean duration of gestation	NA	N	N	↑	Increased 1 day in F1b litters at ≥80 mg/kg/day (may be hormone mediated but not due to estrogen or androgen interactions; may be secondary to systemic toxicity at exaggerated high dose due to mis-dosing)
Gestation index	NA	N	N	N	No effect
Pup sex ratio	NA	N	N	(N)	Increased number of male pups at 80 mg/kg/day in F1a litters; no effect on F1b litters dosed at a higher level; therefore, considered incidental
Uterine histopathology	NA	N	N	N	No effect (weanlings)
Ovarian histopathology	NA	N	N	N	No effect
Androgen pathway –potentially indicative endpoints					
Fertility index	NA	N	N	↓	Decreased (not statistically significantly) in mating to produce F1b litters at ≥80 mg/kg/day
Pup sex ratio	NA	N	N	(N)	Increased number of male pups at 80 mg/kg/day in F1a litters; no effect on F1b litters dosed at higher level; therefore, considered incidental
Testes weight	NA	N	N	N	No effect
Testes histopathology	NA	N	N	N	No effect
Prostate histopathology	NA	N	N	N	No effect (weanlings)
Epididymides histopathology	NA	N	N	N	No effect
Seminal vesicle histopathology	NA	N	N	N	No effect (weanlings)

*Mis-dosing during F1b gestation probably to a dose approximating 100 mg/kg/day.
NA: not applicable; N: no effect; (N): finding but not likely exposure related.
↑Increase
↓Decrease

the thyroid hormone content in blood was measured during a routine biennial health monitoring of the staff (population size not provided), and no attempt was made to correlate thyroid hormone levels with the urine and blood 2,4-D levels. Therefore, although this study does not provide any evidence of thyroid-modulating potential of 2,4-D, it is considered of limited reliability.

Goldner et al. 2010 and Goldner et al. 2013

There are two publications on thyroid disease using data from the Agricultural Health Study (AHS). The first, (Goldner et al. 2010) evaluated exclusively women in the AHS, and the second, (Goldner et al. 2013), evaluated men. The 2010 publication was based on about 16 500 women. This represented about 70% of the overall AHS female cohort. The authors found no association with hypothyroidism and working/living on a farm or using pesticides. They found no statistically significant associations with thyroid disease and 2,4-D.

The 2013 publication included 22 246 men. This group represented only 62% of the overall AHS male cohort because of requirements for complete data on thyroid disease. Contrary to the 2010 Goldner et al. study results, a number of statistically significant odds ratios were reported for various pesticides. These included six herbicides (including 2,4-D), most organochlorines, two insecticides and one carbamate. According to the authors, this is the first epidemiology study to report these associations. With respect to 2,4-D, the authors cite a 2007 paper by Stoker et al. (see Table 11) as supportive evidence of biological plausibility. The latter assertion will be discussed in the thyroid WoE discussion.

The authors list the limitations as "potential for recall bias affecting exposure estimates, reliance on self-reported disease, and possible selection bias due to high dropout rates."

Table 10. Regulatory 2,4-D developmental, subchronic and chronic mammalian studies and evaluation of study quality.

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Developmental studies						
Rodwell 1983	OPP 83-3 Guideline developmental toxicity evaluation in F344 rats. Dosing by oral gavage from GD 6-15.	97.5%	8, 25, 75	*Maternal*: 75 mg/kg/day produced slight toxicity *Developmental*: No adverse effects observed	1	None noted; however note current Guideline requires a longer exposure period; high dose exceeds TSRC
Hoberman 1990	OPP 83-3 Guideline developmental toxicity evaluation in New Zealand white rabbits. Dosing by oral gavage from GD 6 18.	97.5%	10, 30, 90	*Maternal*: High dose produced two abortions *Developmental*: No toxicity observed.	1	None noted; however note current Guideline requires a longer exposure period and more animals per dose level (20 vs 12); high dose exceeds TSRC in rabbits
Subchronic and chronic toxicity studies						
Schulze 1991a	OPP 82-1 guideline subchronic study with F344 rats. Dosing via the diet for 13 weeks.	96.1%	1, 15, 100, 300	Systemic toxicity: 300 mg/kg/day: marked (excessive) systemic toxicity, ↓ body weight (28%), renal toxicity and effects on eyes, hearts, and lungs. 100 mg/kg/day: renal toxicity Endocrine endpoints: 300 mg/kg/day (F): ↓ T3, T4, ↑ thyroid/parathyroid weights, thyroid follicular cell hypertrophy ↓ pituitary weights 100 mg/kg/day (F): ↓ T3 and T4 300 mg/kg/day (M): ↓ T4, ↑ thyroid/parathyroid weights, ↓ testes weight and atrophy; ↓ pituitary weights 100 mg/kg/day (M): ↓ T4; ↑ thyroid/parathyroid weights No histopathological changes: pituitary, epididymides, ovary, uterus, and vagina.	1	High dose excessive; exceeded MTD; doses ≥ 100 mg/kg/day exceed TSRC.
Gorzinski et al. 1991a	OPP 82-1 non-Guideline study with F344 rats. Dosing via the diet for 13 weeks.	97.3%	15, 60, 100, 150	150 mg/kg/day: marked Systemic toxicity in both genders. ~100 mg/kg/day (F): ↓T4 . ~100 mg/kg/day (M): ↓testes weight	2	Not all guideline endpoints evaluated. Doses ≥60 mg/kg/day exceed TSRC.
Gorzinski et al. 1991b	Non-guideline subchronic study with F344 rats. Dosing via the diet for 13 weeks.	100%	15, 60, 100, 150	Systemic toxicity in both genders at ≥~100 mg/kg/day. ≤ 60 mg/kg/day (F): ↓T4. (M): No effects on endocrine relevant endpoints including testes weights	2	Doses ≥60 mg/kg/day exceed TSRC. Very limited endpoints evaluated. Study done primarily to evaluate testes weight change in prior study

(*continued*)

Table 10. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Jeffries et al. 1995	Guideline chronic toxicity/oncogenicity study with F344 rats. Dosing via the diet for 12 or 24 months.	96.45%	5, 75, 150	150 mg/kg/day (F): ↓body weight, renal and other systemic toxicity; ↓secretory material in thyroidal epithelial cells, ↓ ovary weight, ↑ incidence of benign adenomas in pituitary, and ↓ mammary gland hyperplasia. ~75 mg/kg/day: renal tox, ↓T4 levels, ↑ thyroid weights, ↑focal cystic dilatation. 150 mg/kg/day (M): ↑ thyroid weights, ↓testes weights. ~75 mg/kg/day (M): ↓ T4	1	Doses ~75 mg/kg/day exceed TSRC; guideline study; thyroid tissue accountability issue in females at terminal sacrifice.
Schulze 1991b	Guideline B6C3F1 mice subchronic toxicity study. Dosing via the diet for 13 weeks.	96.1%	1, 15, 100, 300	~100 mg/kg/day: ↓ T4 levels at 100 and 300 mg/kg/day, no effect on testes or ovary weights. No histopath. findings: pituitary, adrenal, thyroid, parathyroid, testes, epididymides, ovary, uterus.	1	Doses ~100 mg/kg/day likely exceed TSRC
Stott 1995a	Oncogenicity study B6C3F1 mice (females only). Dosing via diet for 24 months; 12-month interim sacrifice. Guideline when considered in conjunction with Stott 1995b study	96.4%	5, 150, 300	~150 mg/kg/day: systemic toxicity (renal). No histopath. findings: pituitary, adrenal, thyroid, ovary, uterus, vagina, mammary.	1	Doses ~150 mg/kg/day likely exceed TSRC
Stott 1995b	Oncogenicity study with B6C3F1 mice (males only). Dosing via the diet for12 or 24 months. Guideline when considered in conjunction with Stott 1995a study.	96.4%	5, 62.5, 125	No significant exposure-related effects. No histopath. findings: testes, epididymides, prostate, seminal vesicles, adrenal, thyroid, pituitary	1	125 mg/kg/day likely exceeds TSRC
Schulze 1990	Guideline subchronic study with dogs (82-1). Dosing via capsule for 13 weeks.	96.1%	0.3, 1, 3, 10	10 mg/kg/day: Severe systemic toxicity including lethality; ↓ testes weight; testicular atrophy. No effects thyroid: T3, T4, weight and histopath; no effect ovary weight; no histopath findings: ovary, uterus, epididymides, pituitary.	1	Guideline study; 10 mg/kg/day dose exceeds MTD; ~3 mg/kg/day exceeds dog TSRC; immaturity of dogs at study initiation limits ability to detect any exposure-related testes lesions due to very high background incidence in published HCD

(continued)

Table 10. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Dalgard 1993d	Guideline subchronic study with dogs. Dosing via diet for ~13 weeks.	96.7%b	0.5, 1, 3.75, 10 (lowered after 8 weeks to 7.5)	, body weight at mid and high dose, ↑ thyroid weight not considered exposure-related. No clearly dose-related effects on testes or prostate histopath; no effects thyroid histopath.	1	Guideline study; doses ~3.75 mg/kg/day exceed dog TSRC; immaturity of dogs at study initiation limits ability to detect any exposure-related testes and prostate lesions due to very high background incidence in HCD
Dalgard 1993b	Guideline chronic study with dogs. Dosing via diet for 1 year.	96.7%b	1, 5, 10 (lowered after 8 weeks to 7.5)	Body weight ↓ in high dose group. Renal pathology and altered BUN and creatinine levels at ~5 mg/kg/day. No effects on thyroid or testes histopath. Absence of testes findings support that findings in subchronic dog studies were not exposure related or at most may reflect delayed development.	1	Guideline study; doses ~5 mg/kg/day exceed dog TSRC.

With respect to bias of exposure, we know that the AHS participants have adequate recall of what was applied but are less reliable with respect to how often and for how many years. Further, application (i.e. use) is a poor proxy for exposure, because the range of exposures from a single application is highly variable as demonstrated from biomonitoring studies (such as the Farm Family Exposure Study (Alexander et al. 2007) and the AHS biomonitoring study (Thomas et al. 2010)). As a result, the efforts to evaluate exposure-response in the AHS are very limited at best. Reliance on self-reporting is less of a concern with the thyroid outcome, as the outcome is physician-diagnosed, but it may be reflected in the declining participation in the AHS over time. It is also possible that persons with health concerns have selectively participated in Phase II and III, which may bias toward an increase in observed disease rates.

There is a lack of correlation between the genders. The authors reported no association of 2,4-D use and hypothyroidism in women (OR ~0.96; 95% CI 0.8–1.1) and a statistically significant but relatively weak association in men (OR ~1.35; 95% CI 1.04–1.76). If the association for any pesticide exposure was causal for thyroid disease, one would expect to see an association in both men and women. According to the Mayo Clinic, women over 50 are at risk for an underactive thyroid, hypothyroidism http://www.mayoclinic.com/health/hypothyroidism/DS00353/DSECTION=risk-factors. It is unclear why the authors observed associations with several pesticides in men but not in women. It may be due to unintentional bias, such as related to exposure misclassification or to participation.

WoE for potential endocrine pathway interactions

As discussed in the Introduction, the following WoE reflects an assessment of whether results might signal a pathway interaction, the relative weight or rank placed on that parameter for specifically and sensitively flagging a potential interaction, and whether a finding (if any) was made only at a systemically toxic or excessive dose. The WoE tables developed for each pathway hypothesis provide primarily a visual representation that assists in identifying patterns of findings that may indicate a potential endocrine pathway interaction. Endpoints are evaluated for consistency within and between studies.

The following format is used for the WoE table contents addressing each potential endocrine pathway interaction.

- Parameters relevant to and negative for specific potential pathway interactions are indicated in dark gray and marked as "N" for negative.
- Parameters with findings potentially supporting an endocrine-pathway-related finding seen only at high doses exceeding the TSRC and/or excessive doses are indicated in light gray and marked as "O" for over the TSRC.
- Parameters with findings at the limit dose (in the FSTRA, in which TK data were not available to advise dose selection) with an unclear relationship to potential endocrine interactions but showing a test article related response are indicated in light gray and marked "L" for limit dose.

Table 11. Published literature references for mammalian studies and evaluation of study quality.

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Developmental studies						
Bage et al. 1973	Pregnant NMRI mice dosed subcutaneously from GD 6–14	Mixture of formulated products, purity of 2,4-D not defined	50 and 110 (2:1 2,4-D and 2,4,5-T)	Increases in fetal resorptions, cleft palate. Decreased fetal body weights; no malformations characteristic of endocrine modulators	3	Weaknesses: 2,4-D not tested separately (although 2,4,5-T was); purity of 2,4-D not defined; maternal toxicity not evaluated; subcutaneous dosing not relevant for risk assessment; Strength; formulation excipients were added to the control group
Cavieres et al. 2002	ND4 mice dosed orally by gavage from GD 6–15	Formulation (Purity unspecified)	0.01–100	Decreases in litter sizes and purported decreases in implantations	3	Weaknesses: Formulation; purity of 2,4-D unspecified; data discrepancies; lack of appropriate concurrent controls; lack of biological plausibility for reported findings
Charles et al. 2001	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D 2-butoxyethyl ester (BEE) from GD 6–15	2,4-D BEE 95.6%	25, 75, 185 ai (17, 51, 125 ae)	Maternal toxicity at 125 mg/kg/day ae; ↑ skeletal variations at 125 mg/kg/day ae; No dose related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds, e.g. nature of visceral malformations not specified; ae doses ≥51 mg/kg/day exceed TSRC.
	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D 2-ethylhexyl ester (EHE) from GD 6–15	2,4-D EHE 95.0%	14.1, 45.2, 135.7 ai (10, 30, 90 ae)	Maternal toxicity (slight) at 30 mg/kg/day ae; ↓fetal-body weights (90 mg/kg/day ae); ↑ skeletal variations at 30 mg/kg/day ae; No visceral malformations; wavy ribs at 10 and 30 mg/kg/day ae not dose related.	2 (for publication)	Overall high quality study, minor limitations in amount of detail reported due to multiple studies and compounds; e.g. skeletal variations at 30 mg/kg/day ae not shown in Table; Table identifies wavy ribs as malformations; typically classified as deviations (Kimmel et al. 2014); doses >90 mg/kg/day ae exceed TSRC.
	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D isopropylamine (IPE) from GD 6–15	2,4-D IPE 97.1%	12.3, 36.9, 123 ai (10, 30, 100 ae)	Maternal toxicity ≥30 mg/kg/day ae; ↓ fetal-body weights and ↑skeletal var. (100 mg/kg/day ae); No dose related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds;
	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D diethanolamine salt (DEA) from GD 6–15	2,4-D DEA Aqueous based manufacturing concentrate (73.1%)	15, 75, 150 ai (10.2, 50.8, 101.6 ae)	Maternal toxicity ≥50.8 mg/kg/day ae; ↓ fetal-body weights; ↑fetal skeletal variations (101.6 mg/kg/day ae); no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds;
	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D dimethylamine (DMA) from GD 6–15	2,4-D DMA Aqueous based manufacturing concentrate (66.2%)	15, 60.2, 120.4 ai (12.5, 50, 100 ae)	Maternal toxicity >50 mg/kg/day ae; ↓ fetal-body weights and ↑ skeletal changes 100 mg/kg/day ae); No exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds;

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
	OPP 83-3 guideline study. Sprague Dawley rats dosed orally by gavage with 2,4-D isopropylamine (IPA) from GD 6–15	2,4-D IPA Aqueous based manufacturing concentrate (50.2%)	22, 65, 190 ai (17, 51, 150 ae)	Slight maternal toxicity (150 mg/kg/day ae); No exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds
	OPP 83-3 guideline study. Sprague-Dawley rats dosed orally by gavage with 2,4-D triisopropylamine (TIPA) from GD 6–15	2,4-D TIPA Aqueous based manufacturing concentrate (72.2%)	32.5, 100, 325 ai (17, 51, 175 ae)	Maternal toxicity (severe) 175 mg/kg/day ae; ↓ fetal-body weights, ↑skeletal changes and malformations 175 mg/kg/day ae No exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gavage with 2,4-D BEE from GD 7–19	2,4-D BEE 95.6%	15, 45, 110 ai (10, 30, 75 ae)	Severe maternal toxicity at >30 mg/kg/day ae; ↑ resorptions at 30 mg/kg/day ae; no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds; weakness: excessive maternal toxicity at mid and high dose
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gavage with 2,4-D EHE from GD 6–18	2,4-D EHE 95.0%	15.1, 45.2, 113.1 ai (10, 30, 75 ae)	Maternal toxicity at 75 mg/kg/day ae; no exposure related visceral malformations	2 (for publication)	Overall high quality study, minor limitations in amount of detail reported due to multiple studies and compounds
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gavage with 2,4-D DEA from GD 6–18	2,4-D DEA Aqueous based manufacturing concentrate (73.1%)	15, 30, 60 ai (10.2, 20.3, 40 6 ae)	Maternal toxicity and ↑ resorptions and skeletal variations at 40.6 mg/kg/day ae; no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gavage with 2,4-D DMA from GD 6–18	2,4-D DMA Aqueous based manufacturing concentrate (66.2%)	12, 36.1, 108.4 ai (10, 30, 90 ae)	Equivocal maternal toxicity; ↑ deaths at mid dose; ↓ litter size low and high dose; no evidence exposure related; no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds Weaknesses: Equivocal maternal toxicity at high dose; decreased number of litters available for evaluation
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gav age with 2,4 D IPA from GD 7–19	2,4-D IPA Aqueous based manufacturing concentrate (50.2%)	13, 38, 95 ai (10, 30, 75 ae)	Severe maternal toxicity at 75 mg/kg/day ae; no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds; Weakness: excessive maternal toxicity at high dose but sufficient litters for evaluation
	OPP 83-3 guideline study. New Zealand white rabbits dosed orally by gavage with 2,4-D TIPA from GD 7–19	2,4-D TIPA Aqueous based manufacturing concentrate (72.2%)	19, 56, 140 ai (10, 30, 75 ae)	Severe maternal toxicity at 75 mg/kg/day ae; no exposure related visceral malformations	2 (for publication)	Overall high quality study; minor limitations in amount of detail reported due to multiple studies and compounds; Weakness: excessive maternal toxicity at high dose but sufficient litters for evaluation

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Collins & Williams 1971	Syrian golden hamsters dosed orally by gavage from GD 6–10	Three different lots of technical 2,4-D	20–100	Increased incidence of fetal abnormalities at high doses; decreased fetal viability	3	Poor methodology (use of different lots of test material), purity undefined; and reporting deficiencies; additionally hamster is a poor model for developmental toxicity evaluations due to multiple spontaneous malformations
Dinamarca et al. 2007	Pregnant mice dosed with technical and formulated 2,4-D in drinking water from GD 0–9	Purity unspecified	0.01–100	No changes in maternal toxicity, body weight gain, implantation sites, resorptions; no endocrine related effects	2	Purity unspecified for technical and formulated 2,4-D; no analytical confirmation of doses; otherwise study quality seemed adequate; refuted findings of Cavieres et al. (2010)
Duffard et al. 1996	Wistar rats orally exposed to 2,4-D DBE	Formulation; concentration and purity of 2,4-D unspecified	70	Increased serotonin levels	3	Unclear methodologies; insufficient data provided; results not specific to endocrine activity
Pochettino et al. 2013	Prenatal and postnatal 2,4-D exposure in pregnant Wistar rats evaluated at PND 45, 60, and 90	Purity unspecified	70	Evidence of potential oxidative stress in various tissues	3	Source of 2,4-D not identified; purity not stated; single dose; no evidence dose analysis conducted; no evidence potential litter effects were accounted for
Maternal nursing behavior and pup body weight effects						
Stürtz et al. 2010	Wistar rats exposed to 2,4-D via intraperitoneal injection	Purity unspecified	50–100	Decreased pup body weights; detectable 2,4-D residues in stomach, blood, brain, and kidney of breast-fed neonates	3	Irrelevant route of exposure (intraperitoneal); purity not specified; no evidence litter effects accounted for
Stürtz et al. 2006	Wistar rats dosed orally with 2,4-D	98%	15–70	Decreased pup body weight gains; decreased lipid content of milk; altered fatty acid content	3	Unclear methodology; no evidence litter effects accounted for; dose formulation procedures questionable; neither effects nor internal dosimetry consistent with other 2,4-D studies
Stürtz et al. 2008	Wistar rats dosed orally with 2,4-D	98%	15–50	Altered dam-pup interactions; increased catecholamine levels; decreased indolamine and prolactin levels	3	Unclear methodology; no evidence litter effects accounted for; dose formulation procedures questionable; limited assessment of maternal nursing behavior, pup observations inconsistent with typical findings in pups suffering from maternal neglect

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Stürtz et al. 2010	Wistar rats dosed orally or IP with 2,4-D	98%	2.5–70	Decreased pup weight gain; decreases in amount of milk ejected, plasma prolactin and oxytocin altered	3	Unclear methodology; no evidence litter effects accounted for; dose formulation procedures questionable; no bedding material mentioned- absence could stress dams; procedure for evaluation of milk production flawed; incorrect statistical procedures
Male reproductive toxicity						
Lamb et al. 1981a	Male C57BL/6N mice dosed orally via diet mixture of 2,4-D and 2,4,5-T; mated with untreated females	98.5% (2,4-D Purity specified in Lamb et al. 1981c toxicity study)	20 and 40 mg/kg/day 2,4-D	No changes in fertility, sperm number, motility, or morphology	2	2,4-D not tested in isolation, but purity and doses of 2,4-D defined; a negative response reported; provides useful information re potential male repro. toxicity
Lamb et al. 1981b	Male C57BL/6N mice dosed orally via diet mixture of 2,4-D and 2,4,5-T; mated with untreated females	98.5% (2,4-D Purity specified in Lamb et al. 1981c toxicity study)	20 and 40 mg/kg/day 2,4-D	No impact of on reproductive performance of males; no changes in development and survival of fetuses and pups	2	2,4-D not tested in isolation, but a negative response reported and purity and doses of 2,4-D defined; provides useful information regarding potential male repro toxicity
Blakley et al. 1989	CD-1 male mice exposed to a mixture of 2,4-D and picloram via drinking water mated to untreated females	Mixture of chemicals; purity of 2,4-D not specified	84–336	High mortality in males at high concentration; no changes in resorptions, implantations; ↓fetal weight at highest dose; ↑ malformations all doses	3	2,4-D not tested in isolation; types of malformations not reported; small group size
Hassanein 2012	Male Sprague-Dawley rats exposed via gavage	Purity unspecified	30	Congestion of blood vessels in testicles and epididymis after two months; necrosis and sloughing seminiferous tubules, necrobiotic changes in epithelial lining of epididymis	3	Unknown purity; single dose exposure
Kim et al. 2002	Hershberger assay in CD rats	Purity unspecified	50	Increased ventral prostate, Cowper's gland and glans penis weights; in testosterone (T) supplemented phase of assay; authors hypothesize 2,4-D synergizes with T in increasing accessory sex tissue weights or inhibited T metabolizing enzymes' activity	3	T dose reporting discrepant (different doses reported in different sections of paper); only one 2,4-D dose tested; inconsistencies in reported methods; T-supplemented Hershberger phase tests anti-androgenicity; not validated to assess androgenicity; organ weights not similar to weights for the same organs reported by other investigators (note article is in Korean; no translation)

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Oakes et al. 2002a	Male Sprague-Dawley rats dosed orally via gavage	Formulation; 2,4-D concentration and purity unspecified	37.5–150	No effects on plasma testosterone; decreased weight gain and decreased testicular weight at high dose	3	Formulation; concentration and purity unspecified poor study design; small sample size; insufficient reporting of data
Stoker et al. 2007	Male Wistar rats exposed to 2,4-D orally by gavage	Purity unspecified	100 and 200 reported in abstract; per Stoker and Zorilla (2010) book chapter 3 and 30 were also evaluated	Delayed PPS at high dose; decreased ventral prostate weight; T and androstenedione decreased at high dose; no change in LH and prolactin; T3 decreased at both high doses; no effects at 30 mg/kg/day based on Stoker and Zorilla 2010 book chapter	4	Insufficient information provided (only reported as abstract and as limited information in book chapter); doses reported in abstract exceed threshold for renal clearance
Subchronic and chronic toxicity						
Charles et al. 1996a	OPP 82-1 Guideline study; Fischer 344 rats exposed to 2,4-D in the diet for 13 weeks	96.1%	1, 15, 100, 300	Reviewed above with regulatory toxicology studies on 2,4-D based on study report (Schulze 1990 a)	2 (for publication)	Limited detail due to multiple studies in one publication
	OPP 82-1 Guideline study Fischer 344 rats exposed to 2,4-D dimethyl amine (DMA), 2,4-D ethylhexyl amine (EHE) in the diet for 13 weeks	66.2% a.i (95.3% dry wt) 2,4-D DMA	1–300 mg/kg/day a.e. (acid equivalent)	300 mg/kg/day a.e., ↓red cell mass, T3 and T4 levels, ↓ ovary and testes weight. ↑ liver, kidney and thyroid weights and cataracts and retinal degeneration (high-dose females).	2 (for publication)	Limited detail due to multiple studies in one publication
	OPP 82-1 Guideline study Fischer 344 rats exposed to 2,4-D ethylhexyl amine (EHE) in the diet for 13 weeks	95.1% a.i 2,4-D EHE	1–300 mg/kg/day a.e. (acid equivalent)	300 mg/kg/day a.e.: ↓red cell mass, T3 and T4 levels, ↓ ovary and testes weight. ↑ in liver, kidney and thyroid weights and cataracts and retinal degeneration (high-dose females).	2 (for publication)	Limited detail due to multiple studies in one publication
Oakes et al. 2002b	Male Sprague-Dawley rats dosed orally with gavage with a formulation of 2,4-D and picloram	Formulation with multiple active ingredients; 2,4-D concentration and purity not specified	37.5–150 (2,4-D)	No effects on plasma T; ↓ weight gain and ↓ testicular weight at 150 mg/kg/day	3	Formulation; 2,4-D concentration and purity not specified; 2,4-D not tested in isolation, poor study design; small sample size; insufficient reporting of data
Kobal et al. 2000	Wistar rats dosed by gavage for ten days	Formulation (2,4-D in formulation 98% pure, but inerts in formulation unspecified)	11 and 110	↓ serum T4 at 110 mg/kg/day; ↓T3 males day 6 at 110 mg/kg/day Statistically significant differences in the low-dose group (* in females and ↓ in males); examination of the figures suggests that these changes were not biologically significant and may reflect inherent variability of the assays. No other thyroid-related parameters were assessed in this study; the biological significance of the findings cannot be assessed	3	T4 pre-test values showed considerable variance; no data are presented for thyroid hormone measurements, only figures, which show mean values only without confidence limits or standard errors. Formulated commercial material with unknown formulation excipients. High dose exceeds TSRC; other toxicity to animals not monitored. TSH, thyroid weight, and thyroid histopathology not evaluated.

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Charles et al. 1996b	OPP 82-1 subchronic and 83-1 chronic guideline studies in purebred beagle dogs exposed to 2,4-D in diet for 90 days or one year	96.7% 2,4-D	0, 0.5, 1.0, 3.75 and 7.5 in subchronic study; 0, 1, 5 and 10/7.5 in chronic study	Studies (Dalgard 1993a, 1993b) reviewed above with regulatory toxicology studies based on study reports	2 (for publication)	Studies reviewed above based on actual study reports
	OPP 82-1 guideline subchronic study in purebred beagle dogs exposed to 2,4-D dimethylamine salt (DMA) for 90 days	66.7% aqueous solution (93.5% dry weight basis)	0, 1, 3.75 and 7.5 (a.e. basis)	↓ body weight gains at high dose; no effects on thyroids; slightly ↓ testes weight at 7.5 mg/kg/day ae; non-exposure related testicular histopathology; slightly ↑incidence of "juvenile/inactive" prostate at 7.5 mg/kg/day ae	2 (for publication)	Limited details included due to multiple studies presented
	OPP 82-1 guideline subchronic study in purebred beagle dogs exposed to 2,4-D 2-ethylhexyl ester (2-EHE) for 90 days	95.1%	0, 1, 3.75 and 7.5 (a.e. basis)	↓body weight gains at mid and high dose; thyroid weights ↑; not statistically significant or dose related; no thyroid histopath; no effects testes weights; no testicular path.	2 (for publication)	Limited details included due to multiple studies presented
Obidike et al. 2012	West African Dwarf goats dosed via diet	Formulation; purity of 2,4-D not specified	75–125 mg/kg of 720 g/L formulation of 2,4-D	Decreases in testicular sperm reserves and in epididymal sperm reserves; hyperemia and edema of the stroma; decrease in Sertoli cells	3	Non-conventional test system; formulation tested; small number of animals of varying ages obtained from different sources; no assurance control animals representative
Rawlings et al. 1998	Female ewes dosed directly to the rumen via gelatin capsules	Formulation; concentration and purity of 2,4-D not defined	10 mg/kg; 3 times per week	Decreased serum T4; no effect on LH, FSH, P, E2, cortisol, insulin; no changes in histopathology of any organs tested	3	Small sample size; formulation tested; concentration and purity of test material not defined; only one dose tested; direct administration into rumen of questionable relevance
Thyroid mechanism related						
Florsheim & Velcoff 1967	Sprague-Dawley rats on low Iodine diet subcutaneously injected with 2,4-D	Formulation	80	Decrease in serum protein-bound iodine; decrease in thyroid to serum radioiodide ratio; no changes in pituitary TSH concentrations, thyroidal cell height, or thyroid histopathology	"5"	Subcutaneous injection, formulation, unspecified purity of test material; dose likely exceeds TSRC; mechanistic study
Florsheim et al. 1963	Sprague-Dawley rats	Formulation	80	Decreased serum thyroxine concentration; increased brain and liver thyroxine concentration; no change in thyroxine half-life, thyroxine distribution or in thyroid histopathology	"5"	Route of administration not specified (probably subcutaneous injection); small sample size formulation, unspecified purity of test material; mechanistic study; dose likely exceeds TSCR

(*continued*)

Table 11. Continued

Study	Assay and test system	Purity 2,4-D	Concentration range or doses tested (mg/kg/day)	Result	Klimisch score	Weaknesses
Van den Berg et al. 1991	Male WAG/MBL rats exposed to 2,4-DB in diet; 2,4-D *tested in vitro*	"Highest purity available"	0.06 mmol/kg 2,4-DB	Lower plasma T4 levels in vivo following 2,4-D & exposure; decreased T4 binding to serum protein *in vitro* after 2,4-D exposure	"5"	2,4-D not tested in vivo; intraperitoneal route irrelevant; only one concentration tested. In vitro mechanistic study: 2,4-D demonstrated a high (70–100%) competition for binding of T4 to transthyretin at a concentration of 100 uM (22 ug/ml – a concentration well above TSRC).

- Endpoints either not relevant to a specific study type, or not assessed in a particular assay are noted with a "-"and are white.

Note: Many of the findings marked "O" have explanations that rule out attribution to an endocrine mechanism or make this much less likely. These findings are indicated in the table footnotes and discussed in more detail in the text following the tables.

Overall, there were no findings for 2,4-D that are considered clearly positive for a direct endocrine pathway potential interaction *in vitro*, and no *in vivo* findings in mammals in robust studies (Klimisch criteria 1 or 2) relevant to potential endocrine pathway interactions at doses below the TSRC.

WoE evaluation for the estrogen hormonal pathway

As outlined in EPA's WoE guidance (US EPA 2011), generally five Tier 1 EDSP screening studies provide data relevant to assessing whether a compound potentially interacts with the estrogen hormonal pathway. In the case of the 2,4-D EDSP requirements, EPA waived the requirement for the EDSP Tier 1 uterotrophic and female pubertal assays, based on a recent EDSP Tier 2 equivalent EOGRT study (Marty et al. 2010). This study provides similar information to the female pubertal study (OPPTS 890.1450; US EPA 2009h), and is considerably more comprehensive because it includes assessment of reproductive parameters and offspring development. Endpoints highly sensitive to estrogenic effects are included in the EOGRT study; however, this study does not provide the more specific (Rank 1) information on potential estrogenicity provided by the uterotrophic assay.

The Rodwell and Brown two-generation reproductive toxicity study provides additional information, including an assessment of uterine histopathology in immature animals, which, as they are unlikely to be cycling, provide a more sensitive model for assessing estrogenic activity than do adults. The subchronic and chronic regulatory toxicity studies in rats, mice and dogs also provide information on relevant end points and provide a resource for evaluating potential species sensitivity.

Additionally, the EDSP Tier 1 *in vitro* ER binding or activation studies provide information relevant to assessing the potential receptor-mediated estrogenic activity of 2,4-D. The Tier 1 EDSP FSTRA provides information from an aquatic species. Endpoints and results relevant to the estrogen pathway from key regulatory studies, including the EDSP Tier 1 FSTRA and the EOGRT study are summarized in Tables 12–14. The one-generation quail reproductive toxicity study also provides relevant information and is discussed in the text following the tables. High quality *in vitro* and *in vivo* data from the published literature and ToxCast™ screening results are also considered important in developing the WoE and are discussed in the text.

The only type of *in vivo* assay considered to provide a clear rank 1 result for estrogenic pathway activity are: 1) the uterotrophic assay (none available for 2,4-D), in which statistically significant and marked uterine weight increases are

Table 12. Results for 2,4-D from EDSP tier 1 *in vitro* and ecotoxicological assays relevant to potential interaction with the estrogen pathway.

Study*	ER binding	ER transactivation	Fish fecundity	Fish fertility	Nuptial tubercles (M)	Nuptial tubercles (F)	Fish gonad somatic index (M)	Fish gonad somatic index (F)	Fish gonad histopathology (M)	Fish gonad histopathology (F)	VTG (M)	VTG (F)
ERB		–	–	–	–	–	–	–	–	–	–	–
ERTA	–		–	–	–	–	–	–	–	–	–	–
FSTRA	–	–	L↓(3)†									

*ERB: OPPTS 890.1250 ER-binding assay (LeBaron et al. 2011a); ERTA: OPPTS 890.1300 ERα transcriptional activation assay (LeBaron & Kan 2011); FSTRA: OPPTS 890.1350 Fish short-term reproduction assay (Marino et al. 2010)

†Marino et al. 2010 decreased fecundity observed only at high (limit dose) dose; non-specific finding; equivocal relationship to pathway interaction

All studies.

NNo evidence of pathway interaction

LPotential evidence of pathway interaction at limit dose only

-Endpoint not evaluated

↓Decreased relative to control

Endpoint scores in parentheses based on Borgert et al. (2014)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

considered a reliable and predictive endpoint of estrogenicity, and the male VTG endpoint in the FSTRA (available for 2,4-D), which when increased appears closely associated with estrogenicity (Borgert et al. 2014). The VTG endpoint for the FSTRA for 2,4-D was negative. No endpoints or assays at this time are considered to provide unequivocal evidence of anti-estrogenicity by themselves. Thus, although the anti-estrogenic hypothesis was evaluated, the rankings in the following tables were developed primarily according to reliability for assessing estrogenic potential.

In vitro *studies*

There are no *in vitro* data supporting a direct interaction of 2,4-D with the ER. The EDSP *in vitro* ER binding (LeBaron et al. 2011a) and ER transactivation assays (LeBaron & Kan 2011) were negative.

Kojima et al. (2004) examined the ability of 2,4-D to act as either an agonist or an antagonist of the human ERα or human ERß in transiently transfected Chinese hamster ovary cells (CHO K1) that also expressed a luciferase reporter plasmid containing an estrogen-responsive element (ERE). There were no 2,4-D related effects in these assays. Kojima et al. is considered to provide reliable corroborating data regarding the lack of ability of 2,4-D to act *in vitro* as an agonist or antagonist at the ER.

One high quality *in vitro* study examined the ability of reagent-grade 2,4-D to induce the proliferation of an estrogen-responsive cell line, MCF-7 (Lin & Garry 2000). Reagent-grade 2,4-D had no effect on MCF-7 cell proliferation. Commercial grade 2,4-D LV4 and 2,4-D amine caused a proliferative response, but, because the reagent-grade 2,4-D was negative, this response in the commercial grade ingredients was thought to be due to other components present in the commercial formulations. The proliferation of MCF-7 cells can occur via either an estrogen-related pathway or a non-estrogen-related mechanism; additionally the reliability of the test method may be influenced by the derivation of the cell line and culture conditions (Odum et al. 1998; Payne et al. 2000). Therefore, a positive response in this assay alone is not necessarily a reliable indicator of a compound's estrogenic potential.

The ToxCast™ assays lack the details in methods and results required to fully establish validity, but serve to generally confirm the absence of *in vitro* effects on ER-binding or transactivation. As noted previously, a recent analysis by Cox et al. (2014) shows good concordance between the ToxCast™ results and EDSP endpoints indicative for potential ER interaction. ToxCast™ assays showed no ER binding or transactivation potential for 2,4-D.

Ecotoxicological studies

In the EDSP Tier 1 FSTRA (Marino et al. 2010, published in Coady et al. 2013), 2,4-D exposed female fish showed decreased fecundity at the high exposure concentration only (nominal limit dose of 100 mg/L). Secondary sex characteristics were not affected in males or females. There were no effects on male or female VTG, fertility or gonadal histopathology in the FSTRA, and it is considered likely that the high dose effects on fecundity reflect stress or uncharacterized systemic toxicity rather than an effect relating to interaction with the estrogen pathway. Notably, there were no effects signifying increased male VTG, which as noted is considered a rank 1, i.e. relatively specific and sensitive endpoint for assessing potential estrogenicity (Borgert et al. 2014).

Data from an acceptable reproductive toxicity study in quail (Mitchell et al. 2000) show no effects potentially related to an estrogen pathway interaction at any dietary concentration. Ottinger et al. (2002) indicates that female quail show declines in productivity following exposure to estrogenic chemicals; no changes in egg production or hatching success were seen in the 2,4-D quail reproduction study.

There are limited data from high quality ecotoxicological studies from the published literature. Crain et al. 1997 and Spiteri et al. 1999 reported results after dosing alligator's eggs with 2,4-D. Estradiol was used as a positive control. In the first study, on pipping, chorio-allantoic fluid was analyzed and blood from hatchlings (10-days post hatch) was analyzed for estradiol. The sex of hatchlings was also determined. In the second study, which followed a similar exposure regimen

Table 13. Results for 2,4-D from regulatory reproductive/developmental toxicity studies relevant to potential interaction with the estrogen pathway.

Study*	Time to mating	Gestation duration	Gestation index	Mating index	Fertility index	Implantations	Pup sex ratio	Anogenital distance (AGD)	Vaginal opening	Estrous cyclicity	Urogenital malformations	Uterus weight	Uterus histopathology	Ovaries weight	Ovaries histopathology	Vaginal histopathology
EOGRT												O↓ (3)[7]				
Two Gen		O↓ (3)[1]					O (3)[1]					–		–		–
Dev Tox Rat	–	–	–	–	–	–[4]	–[4]	–	–	–		–	–	–	–	–
Dev Tox Rabbit	–	–	–	–	–	–[4]	–[4]	–	–	–		–	–	–	–	–

*EOGRT: F1-extended one-generation reproductive toxicity study (Marty et al. 2010); Two Gen: OPP 83-4 Two-generation reproductive toxicity study (Rodwell & Brown 1985); Dev Tox Rat: OPP 83-3 Rat Developmental toxicity study (Rodwell 1983); Dev Tox Rabbit: OPP 83-3 Rabbit Developmental toxicity study (Haberman 1990).

†Marty et al. 2010 Uterine weight differences were not statistically different from control and attributable to non-exposure related difference from control in stage of estrous cycle at termination; uterine weights were within HCD.

‡Rodwell and Brown 1985 The length of gestation was statistically significantly prolonged (by 1 day) in the production of the F1b pups at ~80 mg/kg/day, compared with controls; may be attributable to the very excessive dose of 2,4-D to the dams during production of the F1b litters; high pup mortality was seen at this excessive dose.

¶80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (increased M pups), compared with the controls. This finding was not repeated in the F1b pups (at a higher dose) and is considered unlikely to be exposure-related because of the lack of consistency.

§The gross and microscopic evaluation of uteri in the F1b PND 28 offspring showed no evidence of imbibed uteri or uterine lining proliferation

|Evaluated but not sensitive; as implantations complete and offspring sex determined prior to initiation of dosing.

#Change in fetal sex ratio present; not attributed to test article as offspring sex determined prior to initiation of dosing; implantations complete prior to dosing; no evidence selective loss of females

All studies

NNo evidence of potential interaction

OFinding over the TSRC

↑Increased relative to control

–Endpoint not evaluated

Endpoint scores in parentheses based on Borgert et al. 2014 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

but incubated eggs at two different temperatures, gonadal histopathology was evaluated. No effects of 2,4-D were found, in contrast to estradiol. In the second study, males and females exposed to 2,4-D were found to hatch at the appropriate temperature for their gender, in contrast to estradiol, in which the hatchlings were 100% female phenotypically, regardless of the incubation temperature. Additionally, gonadal histopathology was not affected at any 2,4-D concentration. These studies were considered valid; the use of a positive control gave the anticipated responses. One reservation was that the amount of 2,4-D penetrating the egg was not determined, therefore predictions based on these studies are limited to the conditions of exposure, i.e. application directly to the egg. These studies show that concentrations of 2,4-D applied to the eggs of alligators did not produce feminization or other evidence of estrogenic effects.

Mammalian studies

There were no effects on females or offspring in the EOGRT study (Marty et al. 2010, published in Marty et al. 2013) supporting either estrogenicity or anti-estrogenicity.

- There were no exposure-related effects on developmental landmarks, including AGD or age at vaginal opening;
- There were no effects on estrous cycle length or estrous cycle pattern, including a lack of either persistent estrus or interrupted cycling, at any dose level;
- There were no exposure related effects on female reproductive indices, including mating, fertility, time to mating, gestation length, pre- and post-implantation loss and corpora lutea number (latter established in a satellite study);
- There were no signs of dystocia in 2,4-D-exposed P1 dams;
- Litter size and pup survival were not affected by 2,4-D;
- There were no biologically significant effects on reproductive organ weights in adults or offspring at any dose of 2,4-D; and
- There was no exposure-related change in reproductive organ histopathology, including ovarian follicle counts.

Thus, there was no pattern suggesting estrogenicity or anti-estrogenicity in female rats exposed to 2,4-D in the EOGRT study.

Although slight increases in uterine weight (rank 3 endpoint in cycling adults) were noted in adult females in the EOGRT at the high-dose level compared to control; these findings are not considered biologically significant because:

- Increases were not statistically significant;
- Findings were in cycling females which have normally variable uterine weights;
- Estrous cyclicity was extensively characterized and was normal;
- The control uterine weights generally fell below HCD;
- The high dose uterine weights were within HCD; and
- There was no correlating histopathology, except for normal estrous cycle related changes.

Table 14. Results for 2,4-D from regulatory toxicity subchronic and chronic toxicity studies relevant to potential interaction with the estrogen pathway.

Study*	Uterus histopathology	Ovaries weight	Ovaries histopathology	Mammary histopathology	Vaginal histopathology
Rat SC (1)		O↑ (2)†		–	
Rat SC (2)		–			–
Rat C		O↓ (3)‡		O(3)¶	
Mouse SC				–	–
Mouse C (F)		–			
Dog SC (1)				–	–
Dog SC (2)					
Dog C					

*Rat SC (1): OPP 82-1 13-week rat subchronic toxicity study (Schulze 1991a); Rat SC (2): OPP 82-1 13-week rat subchronic toxicity study (Gorzinski et al. 1981a); Rat C: OPP 83-5 two year rat chronic/oncogenicity study (Jeffries et al. 1995); Mouse SC: OPP 82-1 13 week mouse subchronic toxicity study (Schulze 1991b); Mouse C(F): OPP 83-2 mouse oncogenicity study females (Stott 1995a); Dog SC (1): OPP 82-1: 13-week dog subchronic toxicity study (Schulze 1991c); Dog SC (2): OPP 82-1 13-week dog subchronic toxicity study (Dalgard 1993a) and Dog C (1): OPP 83-1: dog chronic toxicity study (Dalgard 1993b).

†Schulze 1991a Ovary weight increase at 300 mg/kg/day high dose; no correlating histopathological findings and stage of estrous cycle not controlled at necropsy.

‡Jeffries et al. 1995 Ovary weight decrease at 150 mg/kg/day high dose terminal sacrifice attributable to body weight loss, no correlating histopathological findings.

¶Decreased incidence of mammary gland hyperplasia at 150 mg/kg/day terminal sacrifice; likely attributable to body weight loss

All studies

NNo evidence of potential interaction

OFinding *over* the threshold for renal saturation

↑Increased relative to control

↓Decreased relative to control

–Endpoint not evaluated

Endpoint scores in parentheses based on Borgert et al. 2011 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis, stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

It should be noted that the uterus was evaluated both grossly and histopathologically in F1b PND 28 F344 female rats (Rodwell & Brown 1985). This evaluation provided information on uterine growth and/or stimulation in weanling animals that, while approaching puberty, were unlikely to be cycling based on the time of puberty onset in F344 rats. The absence of cycling makes these young animals less variable and more sensitive to potential estrogenic effects. There were no effects on the uterine histopathology in these animals even at a dose well above the TSRC.

In other regulatory guideline toxicity studies, including a two-generation rat reproductive toxicity study, rat and rabbit developmental toxicity studies in rats, and subchronic and chronic toxicity studies in mice, rats and dogs, few endpoints were observed suggesting either estrogenic or anti estrogenic activity, even at dose levels causing significant systemic toxicity.

One finding in the two-generation reproductive toxicity (Rodwell & Brown 1985) suggesting possible endocrine toxicity (but not necessarily an interaction with the estrogen pathway) was that the length of gestation was statistically significantly prolonged (by 1 day) in the production of the F1b pups at ≥80 mg/kg/day (HDT), compared with controls. (Due to an inadvertent dosing error to the P animals during production of the F1b treatment group, the actual 2,4 D dose for that generation/littering was ≥100 mg/kg/day). Length of gestation is considered a Rank 3 endpoint. Gestation may be prolonged because of difficulties in parturition, hormonal imbalance, delays in implantation or decreased intrauterine growth; other uncharacterized factors may also result in prolonged gestation. The first alternative is unlikely because no evidence of dystocia was reported. Because of the mis-dosing, the high dose was well above the TSRC and significant toxicity was observed, and this delay may simply be due to excessive toxicity. There were no similar findings in the F1a littering (high dose confirmed as 80 mg/kg/day) in the Rodwell and Brown (1985) study or in the Saghir et al. (2008a) range-finding study, in which a similar high and toxic dose did not result in prolonged gestation. At most, the finding of prolonged gestation in the F1b litters provides equivocal evidence of a potentially treatment-related hormonal imbalance resulting from 2,4-D exposure at a dose significantly exceeding the TSRC (and exceeding a classically defined MTD).

With a single exception, higher significance (Rank 2) endpoints in the mammalian regulatory studies showed no effects of 2,4-D suggesting an interaction with the estrogen pathway. The only exception was a change in ovary weight in a subchronic rat study (Schulze 1991a) at a high dose exceeding both the MTD and the TSRC; however, there were no correlating histopathological changes. The stage of the estrous cycle was not controlled in this study and the ovarian weight findings could have been due to chance. No other possibly estrogen pathway-related effects were seen in females in the subchronic rat toxicity study. More robust ovarian assessment, such as in the EOGRT study in which ovarian follicular counts were performed, showed no exposure-related effects.

Two Rank 3 endpoints were affected in the chronic rat study (Jeffries et al. 1995): ovary weight was decreased (with no histopathological correlate) and mammary hyperplasia was decreased, both at the terminal sacrifice at the high excessively toxic dose of 150 mg/kg/day. There was also a decreased incidence of benign adenomas of the *pars distalis* in the pituitary (which is an estrogen-sensitive tumor; Dinse et al. 2010), in females at 150 mg/kg/day. These findings might point toward anti-estrogenicity, but are confounded by the systemic toxicity which included marked body weight

loss. Note that mammary histopathology in the chronic rat study is ranked a 3 for sensitivity (as opposed to 2 for this endpoint in subchronic studies) because the high background incidence of mammary tumors in the F344 rat strain is a potential confounder.

The mouse subchronic (Schulze 1991b) and chronic (Stott 1995a) studies showed no effects on uterine or ovarian histopathology. The dog subchronic (Schulze 1990 and Dalgard 1993a) and chronic studies (Dalgard 1993b) similarly showed no effects on ovary weights, or on uterine, vaginal or mammary gland histopathology.

As noted previously, estrogens could alter male reproductive system endpoints including testes weight, sperm development or histopathology. The Marty et al. (2010) EOGRT study showed no testicular weight or histopathology findings attributable to a potential endocrine pathway interaction. Sperm parameters including testicular and epididymal sperm counts, motility and morphology were not altered in the Marty et al. (2010) study. Testicular weights and histopathology in the Rodwell and Brown (1984) two-generation reproductive toxicity study showed no exposure-related findings. Findings in these studies in males and on male reproductive tissues in the subchronic and chronic toxicity studies will be discussed in detail in the assessment of potential interactions with the androgen pathway.

Other mammalian studies relevant to potential interactions with the estrogen pathway

In a study by Dinamarca et al. (2007), ICR/Jcl mice were mated, and subsequently administered 2,4 D as a "pure compound" (purity unspecified) or as a commercially available formulation available in Chile (unspecified) in drinking water at concentrations providing mg/kg/day doses of 0, 0.01, 0.10 or 100 mg/kg/day from gestational day (GD) 0–9. The dose spacing in this study was designed to address the low dose hypothesis proposed by Cavieres et al. (2002). (The Cavieres et al. research is summarized in Supplementary Appendix VII.) Maternal toxicity was evaluated. Mice were bled at GD 9 for biochemical evaluations and cesarean-sectioned. Ovaries were evaluated for numbers of corpora lutea and uterine horns were evaluated for number of implantation sites, resorptions and live embryos. There were no signs of maternal toxicity nor differences in body weight gain between the dosed groups and the control. Numbers of corpora lutea, implantation sites, resorptions and live embryos were similar between the dose groups and control. The Dinamarca et al. study demonstrated that the finding of decreased implantations reported in mice exposed to 2,4-D by Cavieres et al. (2002) could not be replicated, even with an exposure period correctly designed to explore this possibility. The only limitations we identified in the Dinamarca et al. study was the fail ure to identify the specific purity of the "pure" test substance, and the lack of confirmatory dose analyses.

Regulatory developmental toxicity studies in rats and rabbits on various esters, amines and salts of 2,4-D summarized by Charles et al. (2001) do not predict any estrogenic activity. This publication is considered Klimisch criteria 2 based on relative absence of detail in reporting because of the large number of studies covered. The individual studies were guideline compliant, with the exception of the rabbit developmental toxicity study of 2,4-D DMA, which had a reduced number of litters available for evaluation. However, developmental toxicity was considered to be adequately characterized in this study. None of the rat or rabbit developmental toxicity studies showed any effects on maintenance of pregnancy, or urogenital malformations of the type that may signify endocrine modulating activity.

Another article by Charles et al. (1996a) presents data from several rat subchronic toxicity studies conducted with 2,4-D DMA, or 2,4-D 2-ethylhexyl ester (2-EHE), and with the 2,4-D acid study by Schulze (1991a) discussed above. This publication is considered Klimisch criteria 2 based on relative absence of detail in reporting because of the large number of studies covered. The studies were GLP guideline studies conducted to satisfy US EPA regulatory testing requirements. Fischer 344 rats (10/sex/dose group) were dosed in the diet with target doses of 0, 1, 15, 100 and 300 mg/kg/day (expressed as acid equivalent doses) for 90 days. Endocrine endpoints relevant to potential interactions with the estrogen pathway included: ovary organ weight, and mammary gland, ovary, and uterine histopathological evaluations. There was no evidence of potential interaction with the estrogen pathway in these studies.

A third article by Charles et al. (1996b) presents data from dog subchronic toxicity studies conducted with 2,4-D DMA, or 2,4-D 2-EHE. It also includes results from subchronic and chronic dog studies on 2,4-D acid (Dalgard 1993a, 1993b), which were reviewed based on the study reports. This publication is considered Klimisch criteria 2 based on relative absence of detail in reporting because of the number of stud ies covered. These studies were GLP guideline studies conducted to satisfy US EPA testing requirements. Beagle dogs (4/sex/dose group) were dosed in the diet with target doses of 0, 1.0, 3.75 and 7.5 mg/kg/day (expressed as acid equivalent doses).

Endocrine endpoints evaluated in these subchronic and chronic studies most relevant to the estrogen pathway included ovary weights, and histopathological evaluations of mammary gland, ovary and uterus. Clinical signs, body weight, feed consumption, clinical pathology and evaluation of standard target organ histopathology were also performed in these studies. There were no dose-related findings in possible estrogen pathway-related female endpoints, even though the high dose caused marked systemic toxicity.

In conclusion, the EDSP Tier 1 assays of 2,4-D considered relevant to the estrogen hormonal pathway, and the Marty et al. (2010) Tier 2 EDSP EOGRT study were judged to be quality studies. Further, the results were considered reliable for assessing the potential interaction with the estrogen pathway for 2,4-D, supplemented by information in other regulatory studies of 2,4-D as well as the high-quality studies available in the published literature. Based on a WoE evaluation of the available data, including the absence of potentially estrogenic or anti-estrogenic exposure-related findings in the Marty et al. (2010) EOGRT study, the lack of evidence for potential estrogen pathway interactions predicted by the other regulatory mammalian toxicity studies at doses below

the TSRC, the weak and non-specific response in the FSTRA, the absence of any adverse effects in the quail dietary reproductive toxicity study, the absence of adverse effects in high quality studies in the published literature and the negative Tier 1 EDSP *in vitro* ER binding and ER transactivation assays as well as negative ToxCast™ and other high quality *in vitro* screening data relevant to the estrogen pathway, it is concluded that 2,4-D does not show evidence for direct interaction with the estrogen pathway.

WoE evaluation for the androgen hormonal pathway

Two EDSP Tier 1 screening assays, including the AR binding assay (LeBaron et al. 2011b, published in Coady et al. 2014) and the FSTRA (Marino et al. 2010, published in Coady et al. 2013), provide data relevant to assessing whether 2,4-D potentially interacts with the androgen hormonal pathway. No Hershberger or male pubertal assays were required for 2,4-D because of the recently completed EOGRT study Marty et al. (2010) (published in Marty et al. 2013). The EOGRT study provides information on all endpoints included in the male pubertal assay, with the exception of serum testosterone levels, and provides additional endpoints sensitive to androgen deficiency not assessed in the pubertal study including AGD and nipple retention in males. Endpoints considered most relevant for assessing potential interactions with the androgen pathway are findings from the Marino et al. (2010) Tier 1 EDSP FSTRA, the Marty et al. (2010) Tier 2 EOGRT study, the Rodwell and Brown (1985) two-generation reproductive toxicity study, the Rodwell (1983) developmental toxicity study, the Hoberman (1990) developmental toxicity study. These studies, along with the subchronic and chronic toxicity studies are summarized below in Tables 15–17.

There were no findings for 2,4-D that suggest a potential androgen pathway interaction at doses less than the TSRC in any mammalian species, and only very limited *in vitro* data suggesting a direct interaction with this pathway is possible. The ranking used in the following tables reflects an assessment of the validity of the endpoints for assessing anti-androgenicity.

As noted previously, the WoE for each pathway depends on multiple other factors besides the rankings, and the rankings themselves may be modified based on the strength of a particular response. The WoE may also be influenced by context, e.g. decreased secondary sex characteristics in male fish is probably a stronger signal for potential anti-androgenicity than for estrogenicity, but may be due to other toxicity or to effects on steroidogenesis or interaction with the HPG axis rather than due specifically to potential anti-androgenicity. In contrast, an increase in nuptial tubercles in female fish appears closely associated with androgenicity and is considered a rank 1 endpoint for assessing the potential interaction with the AR (Borgert et al. 2014). Additionally, typical variance in a parameter in an assay should be considered in ranking that parameter.

More confidence in a potential anti-androgenic pathway interaction would come from corroborative findings in the *in vitro* AR binding and transactivation assays. Findings such as markedly delayed (2–3 day) preputial separation in a reproductive toxicity assay, urogenital anomalies in fetuses or pups exposed *in utero* in developmental or reproductive toxicity studies *and* alterations in male gonad histopathology in repeat dose studies and diminished male secondary sex characteristics in the FSTRA would strongly signal anti-androgenicity, particularly if the *in vivo* findings occurred at not-otherwise systemically toxic concentrations. Importantly, no such corroborative evidence of anti-androgenicity was observed across the 2,4-D studies.

In vitro *studies*

The *in vitro* EDSP Tier 1 AR binding assay (LeBaron et al. 2011b, published in Coady et al. 2013) for 2,4-D was negative.

In a high quality published *in vitro* study, Kojima et al. (2004) examined the ability of 2,4 D to act as either an agonist or an antagonist of the human AR in transiently transfected CHO K1 that also expressed a luciferase reporter plasmid containing an androgen responsive element (ARE) over a range of test compound concentrations (10^{-8}–10^{-5} M). Kojima et al. found no effects associated with 2,4-D exposure and is considered to provide reliable data regarding the lack of ability of 2,4-D to act as an agonist or antagonist at the AR *in vitro*.

Sun et al. (2012) reported the results of luciferase reporter gene assays to measure the effects of 2,4-D on AR in Vero cells, derived from African green monkey kidney epithelium, which do not express the AR endogenously. Transfected cells were exposed to a range of concentrations of 2,4-D based on the maximum contaminant level [allowed] (MCL) in Chinese drinking water, i.e. cells were treated with 0.003, 0.03, 0.3 and 3.0 mg/L 2,4-D. There was no detectable androgenic or anti-androgenic activity at any of the 2,4-D concentrations tested; however, at the high concentration defined as 100x the MCL, 2,4-D was reported to enhance the effects of testosterone in the AR antagonist assay. The rationale for dose selection was questionable in this study; the effect was seen only at a high concentration far exceeding potential human exposure (Aylward & Hays 2008) although it falls within the linear TK range in rats. Further, the biological basis of this assay and relevance of this finding is questionable as the test was specifically designed to measure anti-androgenic activity rather than potentiation or androgenic activity.

The ToxCast™ assay battery developed and run under the auspices of the US EPA showed no evidence of interaction with the AR in either a cell-based or cell-free system. ToxCast™ assays showed no AR binding or transactivation potential for 2,4-D.

Similar to findings for estrogen pathway-related activity, an analysis by Rotroff et al. (2013) and a separate analysis by Cox et al. (2014) showed relatively good concordance between the ToxCast™ results and EDSP Tier 1 endpoints indicative for potential AR interaction, including Hershberger assay results.

Ecotoxicological studies

In the FSTRA (Marino et al. 2010, published in Coady et al. 2013), there were no specific effects on male fish suggesting

Table 15. Results for 2,4-D from EDSP tier 1 *in vitro* and ecotoxicological assays relevant to potential interaction with the androgen pathway.

Study*	AR binding	Fish behavior (sex-linked)†	Fish fertility	Fish fecundity	Fish nuptial tubercles (M)	Fish nuptial tubercles (F)	Fish GSI (M)	Fish GSI (F)	Fish gonad histopathology (M)	Fish gonad histopathology (F)	VTG (M)	VTG (F)
ARB	[redacted]	–	–	–	–	–	–	–	–	–	–	–
FSTRA	–	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*ARB: OPPTS 890.1250 AR-binding assay (LeBaron et al. 2011b); FSTRA: OPPTS 890.1350 fish short-term reproduction assay (Marino et al. 2010).
†Based on impression of typical male nest-guarding behavior which may be associated with testosterone concentrations (O'Connor et al. 2011); no detailed assessment conducted
All studies
NNo evidence of potential interaction
–Endpoint not evaluated
Endpoint scores in parentheses based on Borgert et al. (2014) (modified in some cases based on context or strength of response)
(1) specific and sensitive to the hypothesis
(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data
(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

an androgen pathway interaction. There were no effects on secondary sex characteristics in female fish (occurrence/increase nuptial tubercles in female fathead minnows is a Rank 1 endpoint for androgenicity), and no effects on VTG or gonadal histopathology in either males or females.

A reproductive toxicity study in quail (Mitchell et al. 2000) did not show any findings potentially associated with interaction with the androgen pathway.

There are limited data from high quality ecotoxicological studies from the published literature. Crain et al. (1997, 1999) and Spiteri et al. (1999) reported results after dosing alligator's eggs with 2,4-D. Estradiol was used as a positive control for estrogen pathway mediated effects; there was no positive control for androgen mediated effects. In the first study, at pipping, chorio-allantoic fluid was analyzed and blood from hatchlings (10 days post hatch) was analyzed for testosterone. The sex of hatchlings was also determined. The 1999 study included evaluation of testicular histopathology. Gonadal histopathology was evaluated in a study by Spiteri et al. (1999) which followed a similar exposure regimen, but incubated eggs at two different temperatures. (Incubation temperature normally determines sex of alligator eggs.) There were no effects of 2,4-D on testosterone concentrations or on testicular histopathology following 2,4-D exposure. The studies were considered valid (Klimisch 2). One reservation was that the amount of 2,4-D penetrating the egg was not determined. Consequently, predictions based on these studies are limited to the conditions of exposure, i.e. application directly to the egg. These studies provide support that 2,4-D applied to alligator eggs at concentrations up to 14 ppm do not alter testosterone concentrations or affect male reproductive system histopathology in the exposed offspring.

Mammalian studies

There were no effects in the Marty et al. (2010) EOGRT study (published in Marty et al. 2013) considered to reflect an androgenic or anti-androgenic mode of action. Although there were several study findings that could support potential anti-androgenicity, these were found either not exposure-related, not replicated across generations, or were attributable to other factors such as randomization artifacts.

Preputial separation was slightly delayed (1.6 days) in the F1 males at 800 ppm, which was attributed to decreased growth during lactation and post-weaning. Although preputial separation delay is a Rank 2 finding for anti-androgenicity, the magnitude of the effect was very slight. Body weight at the time of puberty onset was similar in 800 ppm males and controls, despite the delay in onset for the high dose group. These data indicate that 800 ppm 2,4-D had an effect on the rate of growth in peri-pubescent male rats. The magnitude of the delay in preputial separation was consistent with reductions in body weight as demonstrated by a feed restriction study (Marty et al. 2003), and supports that the decreased growth, as opposed to anti-androgenicity, was responsible for the delay in preputial separation in the 2,4-D EOGRT study.

In P1 adult males, decreased seminal vesicle and prostate weights (without corresponding histopathological changes) were seen at $\geq$300 ppm; prostate weights were not statistically different from control. The control weights for the seminal vesicles and prostate were atypically high compared to HCD, and the high dose findings were within HCD. Two high dose males showed testicular atrophy (within HCD). None of these Rank 2 findings were reproduced in the F1 generation, which had higher and longer duration exposures. As noted previously, based on the absence of dose-related increased implantation loss or fetal deaths, there was no evidence that a sensitive sub-population was removed.

Decreased testis weights in 600-ppm PND 22 F1 weanlings were attributed to decreased body weights. A previous feed restriction study with untreated rats showed weanling testes weights alter with decreased body weight (Carney et al. 2004). In contrast, testes weights are conserved in the presence of similar/modest body weight decrements in adult male rats (Chapin et al. 1993). In the EOGRT study, there were no histopathological findings in the testes, and no effects on testis weights or histopathology were seen in adult F1 males.

Other endpoints sensitive to anti-androgenicity, including AGD and nipple retention in F1 offspring, were not altered. As noted in Marty et al. 2013, "These endpoints are considered highly sensitive to altered androgen status (Clark 1999; McIntyre et al. 2001; Wolf et al. 2002; Hotchkiss et al. 2004)."

Other androgen-sensitive endpoints examined in F1 generation adults, including sperm parameters, male reproductive

Table 16. Results for 2,4-D from regulatory reproductive and developmental toxicity studies relevant to potential interaction with the androgen pathway.

Study*	Mating indices	Fertility indices	Anogenital distance (AGD)	Nipple retention	Pup sex ratio	Developmental (e.g. urogenital) malformations	Preputial separation	Sperm parameters	Testes weight	Testes histopathology	Epididymides weight	Epididymides histopathology	Seminal vesicles weight	Seminal vesicles histopathology	Prostate weight	Prostate histopathology
EOGRT	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	--	O(2)†	[redacted]	O (2)‡	O(2)¶	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Two Gen Repro	--	--	--	--	Ø (3)§	[redacted]	--	--	--	[redacted]	--	[redacted]	--	[redacted]	--	N(2)
Dev Tox Rat	--	--	--	--	↓	[redacted]	--	--	--	--	--	--	--	--	--	--
Dev Tox Rabbit	--	--	--	-**	↓	N (2)	--	--	--	--	--	--	--	--	--	--

*EOGRT: Tier 2 equivalent F1-extended one-generation reproductive toxicity study (Marty et al. 2010); Two Gen: OPPTS 83-4 Two-generation reproductive toxicity study (Rodwell & Brown 1985); Dev Tox Rat: 83-3 developmental toxicity study in rats (Rodwell 1983); Dev Tox Rabbit: OPP 83-3 developmental toxicity study in rabbits (Hoberman 1990).
†Marty et al. 2010 Slight delay in preputial separation considered associated with decreased pup weight post weaning and not endocrine related.
‡Decreased testes weight F1 pups at weaning, attributable to decreased body weight, not apparent in F1 adults.
¶Testicular atrophy in 2 high-dose P adults, considered spontaneous, not seen in F1 dosed adults and incidence within HCD.
§Seminal vesicle and prostate weights decreased compared to control at ≥300 ppm in P1; not statistically significant, within HCD and controls below HCD; conclusion no exposure-related effect.
||Rodwell and Brown 1985 Pup sex ratio statistically significantly different in F1a high dose only; but not in F1b which received a higher dose (due to mis-dosing) during gestation; altered ratio considered incidental.
#Evaluated but not sensitive; as offspring sex determined prior to initiation of dosing.
**Change in fetal sex ratio present; not attributed to test article as offspring sex determined prior to initiation of dosing; no evidence selective loss of females
All studies:
NNo evidence of potential interaction
OFinding over the threshold for renal saturation
↓Decreased relative to control
–Endpoint not evaluated
Endpoint scores in parentheses based on Borgert et al. 2014 (modified in some cases based on context or strength of response)
(1) specific and sensitive to the hypothesis
(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data
(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

organ weights, and testicular and accessory sex gland weights or histopathology, were not altered.

Further, although this has not been tested formally, the lower sensitivity to males to systemic toxicity from 2,4-D exposure compared to females in the EOGRT study correlates with the higher predicted androgen-mediated expression of OAT-1 in males (Ljubojevic et al. 2004). This sex-specific difference would not be expected to be prominent if 2,4-D had potent anti-androgenic activity.

There were relatively a few relevant parameters evaluated in the Rodwell and Brown (1985) two-generation study. Mating and fertility were not affected, nor were testicular weights and histopathology, or histopathology of the epididymides, seminal vesicles and prostate at doses up to 80 mg/kg/day.

A single finding in the Rodwell and Brown (1985) two-generation study potentially showing an androgen pathway interaction was that the 80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (109 males and 71 females), compared with the controls. This finding was not repeated in the F1b pups (at a higher dose, due to mis-dosing during gestation) and is considered unlikely to be exposure related because of the lack of consistency. Additionally, there were no parallel exposure related effects on the sex ratio in the EOGRT study for 2,4-D (Marty et al. 2010), or in the range-finding for the latter study (Saghir et al. 2008a) at equivalent or higher concentrations to those tested by Rodwell and Brown (1985). Therefore, the finding is considered not likely to be exposure related.

There were no changes in fetal sex ratio noted in the developmental study in rats (Rodwell 1983). This is not considered a relevant endpoint in this study design because the genotypic sex of offspring is determined prior to the time during gestation that dosing was initiated. The developmental toxicity study in rabbits (Hoberman 1990) showed an altered sex ratio (more males than females) at the high dose. This finding is not considered exposure-related because dosing in this study was initiated shortly after implantation, at a time when genotypic sex of offspring has already been determined. There was no exposure-related selective loss of female offspring *in utero* based on a lack of effects on post-implantation loss or on litter size. Neither developmental study showed exposure-related urogenital visceral malformations which could reflect anti-androgenicity.

The Schulze (1991a) subchronic rat study showed flaccid testes and testicular atrophy at the very high and systemically toxic dose of 300 mg/kg/day, far exceeding the TSRC and MTD. Decreased testes weights were seen in a subchronic rat toxicity study at doses ≥100 mg/kg/day, also exceeding the TSRC (Gorzinski et al. 1981a); however, a follow up study by the same authors (Gorzinski et al. 1981b) failed to replicate this finding. Neither of the two Gorzinski et al. studies showed exposure-related histopathological lesions in the testes. Decreased testes weight was also seen at the excessively systemically toxic high dose (150 mg/kg/day) in the chronic rat study (Jeffries et al. 1995). Testicular atrophy was noted in 2/10 animals at the interim sacrifice in this study; however, no exposure-related testicular lesions were evident at the terminal sacrifice. The high-dose testes findings in the rat

Table 17. Results for 2,4-D from regulatory subchronic and chronic toxicity studies in rat, mouse and dog relevant to potential interaction with the androgen pathway.

Study*	Testes weight	Testes histopathology	Epididymides weight	Epididymides histopathology	Seminal vesicles histopathology	Prostate histopathology	Coagulating glands histopathology
Rat SC (1)	O↓ (2)†	O (2)†			–	–	–
Rat SC (2)	O↓ (2)†	N(2)	–				
Rat SC (3)	N(2)	N(2)	–	–	–	–	–
Rat C	O↓ (2)‡	O (2)‡	–				–
Mouse SC			–		–	–	–
Mouse C (M)			–				
Dog SC (1)	O↓ (3)§	O (3)‖	O↓(3)#		–	–	–
Dog SC (2)	O↓ (3)**		–		–		–
Dog C			–		–		–

*Rat SC (1): 13 week rat subchronic toxicity study (Schulze 1991a); SC (2): 13 week rat subchronic toxicity study 82-1 (Gorzinski et al. 1981a); SC (3): 13 week rat subchronic toxicity study (Gorzinski et al. 1981b); Rat C. OPP 83-5. Two year rat chronic/oncogenicity study (Jeffries et al. 1995); Mouse SC: OPP 82-1 13 week mouse subchronic toxicity study (Schulze 1991b); Mouse C (M): OPP 83-2 mouse oncogenicity study (males) (Stott 1995b); Dog SC (1): OPP 82-1: 13-week dog subchronic toxicity study (Schulze 1990), Dog SC (2): OPP 82 1: 13-week dog subchronic toxicity study (Dalgard 1993a) and Dog C. OPP 83-1: dog chronic toxicity study (Dalgard 1993b).

†Schulze 1991a and Gorzinski et al. 1981a Testes weight: decrease at ≥100 mg/kg/day; atrophy at 300 mg/kg/day (doses far exceeded TSRC and systemically toxic).

‡Jeffries et al. 1995 Atrophy in 2/10 animals at 150 mg/kg/day in interim sac; not evident at termination (dose far exceeded TSRC and systemically toxic).

¶Stott 1995b Relative weight increased but not absolute weight at 24 month only; attributed to body weight decrease.

§Schulze 1990 Testes weight decreased at 10 mg/kg/day (dose exceeded TSRC and MTD); juvenile animals tested.

‖Hypospermia in 2 of the 10 mg/kg/day dose group surviving dogs, equivocal relationship to exposure (dose exceeded MTD) and juvenile animals tested.

#Epididymal weight taken with testes weight; see comments on testes; no histopathological correlates in epididymides.

**Dalgard 1993a Testes weight decreased at 7.5 mg/kg/day dose (dose systemically toxic and exceeded TSRC); juvenile animals tested

††"Juvenile" prostate noted in all groups; not exposure-related

All studies:

NNo evidence of potential interaction

OFinding over the threshold for renal saturation

↑Increased relative to control

↓Decreased relative to control

–Endpoint not evaluated

Endpoint scores in parenthesis based on Borgert et al. 2014 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

studies could hypothetically be associated with anti-estrogenicity or with androgenicity; they could as readily hypothetically reflect changes in steroidogenesis caused by the severe systemic toxicity at the excessive dose levels which far exceed the TSRC, or simply reflect direct non-endocrine mediated target organ toxicity, such as that induced by oxidative stress. Given that there is no other potential evidence of androgenicity or anti-androgenicity in these studies, alternative mechanisms present a more likely hypothesis.

Decreased testicular weights were noted in two subchronic dog studies (Schulze 1990; Dalgard 1993a) at high (lethal or close to lethal) dose levels clearly exceeding the TSRC (and classically defined MTD) in that species. The Schulze (1990) subchronic dog study also showed an increased incidence of testicular lesions (giant cells and hypospermatogenesis) albeit at a lethal dose. Evaluation of the two 2,4-D dog subchronic studies together, which were conducted at the same laboratory, however, shows a high combined control incidence of these lesions. A high historical control incidence has also been reported for these findings in the histopathological literature, particularly for juvenile dogs. Immature dogs (less than 9 months of age) have been reported to have control incidences as high as 75% of both decreased testes weight and hypospermia (Goedken et al. 2008). In the Goedken et al. analyzes of data from a large population of control dogs, atrophic/hypoplastic tubules in the testes were seen in 26.3% of all dogs, with 25–40% of dogs under 12 months old showing this finding. Evaluation of body weights, and control testicular and prostate histopathology (detailed in Supplementary Appendix IV) demonstrates the dogs in these subchronic studies of 2,4-D were on the lower end of the stated age range of 4–6 months old at study initiation and clearly juvenile (less than 9 months old) at terminal sacrifice. Additionally, the high incidence of finding of "juvenile prostate" in the Dalgard 1993a study (prostate was not evaluated in the Schulze (1990) study) supports that the dogs were immature. It seems likely that the decreased testes weights in both subchronic studies and histopathological findings in the testes of the dogs in the Schulze (1990) study are an artifact related to the young age of these animals, or at most, represent delayed development caused by the extremely toxic high dose. Supporting this possibility, a chronic study in dogs (Dalgard 1993b) showed no exposure-related effects on testes weights or histopathology following a one-year exposure to 2,4-D at a high dose level (10 reduced to 7.5 mg/kg/day), generally equivalent to the high dose in the prior subchronic studies. Reversibility of an exposure-related testicular finding such as atrophy in a continuous exposure situation is very unlikely; so the absence of effects in the dog chronic study supports that the findings in the subchronic dog studies were not exposure-related.

Published mammalian studies relevant to the androgen pathway

Two studies by Lamb et al. (1981a, 1981b) evaluated potential male reproductive effects. In both studies, male C57BL/6N

mice were dosed for 8 weeks with combinations of organochlorine chemicals including 2,4-D. Approximate exposures to 2,4-D were 40 mg/kg/day and 20 mg/kg/day (in combination with varied concentrations of the other compounds). The high dose of 2,4-D is considered likely to approximate the TSRC (rats and mice have similar expression of the OAT-1 transporter (Buist & Klaassen 2004)), so the 2,4-D dose was appropriate and not limited by the toxicity of the other mixture components. Controls received untreated diet. Lamb et al. (1981a, 1981b) included evaluation of male fertility (mating the dosed males with untreated females), and sperm number, motility and morphology. Females were either cesarean sectioned at gestation day 18 for fetal evaluation or allowed to deliver and rear their pups until PND 21 for evaluation of offspring birth weight and viability. There was no effect on male fertility, sperm parameters or reproductive performance of the dosed males; development and survival of fetuses and pups in the dosed groups were similar to that of the control mice. The study provided no evidence of male-mediated reproductive toxicity or of endocrine disrupting activity of 2,4 D in these mixtures of chemicals and provide data supporting that there is no evidence of potential androgen pathway interactions for 2,4 D. The studies are high quality; the primary weakness in these studies, as it relates to 2,4-D specifically, is that they test only mixtures.

Regulatory developmental toxicity studies in rats and rabbits on various esters, amines and salts of 2,4-D summarized by Charles et al. (2001) do not predict any androgenic activity. None of the rat or rabbit developmental toxicity studies showed any urogenital malformations of the type that may signify endocrine modulating activity.

An article by Charles et al. (1996a) presents data from rat subchronic toxicity studies conducted with 2,4-D DMA, or 2,4-D 2-EHE as well as with the acid (discussed above (Schulze 1991a)). These studies were GLP guideline studies conducted to satisfy US EPA testing requirements. Fischer 344 rats (10/sex/dose group) were dosed in the diet with target doses of 0, 1, 15, 100 and 300 mg/kg/day (expressed as acid equivalent doses) for 90 days. Endocrine endpoints relevant to potential interactions with the androgen pathway included: testes weight; and epididymides, prostate and testes, histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology and evaluation of standard target organ histopathology were also assessed in these studies. Findings were similar to those reported for 2,4-D acid, and all occurred at a very high and severely systemically toxic dose level. Relative testes weights were decreased at 300 mg/kg/day (acid equivalent), and testicular atrophy was noted at the same dose level. This dose far exceeds the TSRC; and the marked systemic toxicity could have contributed to the testes findings.

A second article by Charles et al. (1996b) presents data from dog subchronic toxicity studies conducted with 2,4-D DMA or 2,4-D 2-EHE (and also included one of the dog subchronic studies (Dalgard 1993a) and the dog chronic toxicity study (Dalgard 1993b) on 2,4-D acid discussed above). These studies were GLP guideline studies conducted to satisfy US EPA testing requirements. Beagle dogs (4/sex/dose group) were dosed in the diet with target doses of 0, 1.0, 3.75 and 7.5 mg/kg/day (expressed as acid equivalent doses). Endocrine endpoints evaluated in these studies possibly relevant to the androgen pathway included testes weight; and epididymides, prostate and testes histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology and evaluation of standard target organ histopathology were also performed in these studies.

There were two findings in these studies that potentially could be related to androgen pathway modulation, although, similar to the hypothesis for the subchronic dog findings with 2,4 D acid, they more likely reflect the immature age of the dogs. Relative testes weights were decreased in the mid dose but not high dose of both the 2,4-D DMA and 2,4-D EHE subchronic studies. (In the subchronic study of 2,4-D acid (Dalgard 1993a) reviewed above, testes weight decreases were seen at the high dose). The absence of dose response in the 2,4-D DMA and 2,4-D EHE studies supports that the finding in the 2,4-D acid study was related to the immature age of the test animals, as discussed previously, and occurred by chance. Additionally, there were no exposure-related histopathological lesions in the testes in the 2,4-D DMA and EHE subchronic studies.

The second finding in the dog studies reported by Charles et al. (1996b) is that inactive/juvenile prostates were noted in several high-dose males in the subchronic studies of 2,4D DMA and EHE; the authors concluded this finding is likely to be related to delayed development from poor nutrition. We consider it likely that this finding also (or possibly primarily) reflects the immature age of the test animals. First, this finding was not made in the chronic 2,4-D acid dog toxicity study which tested a similar high dose on an acid equivalent basis. Dogs at study initiation in the subchronic 2,4-D DMA and EHE studies were relatively young (4–6 months old at study initiation according to Charles et al. (1996b)); review of body weight data suggests many of the dogs were on the low end of this age range. In the subchronic 2,4-D acid dog study (Dalgard 1993a), the incidence of the juvenile/inactive prostate finding shows a clearly non-dose related pattern, including the presence of this finding in control dogs. As discussed above in the context of the 2,4-D acid dog studies, a very high incidence of decreased testes weights and the same testicular lesions observed in these studies has been reported in young control dogs; an increased incidence of juvenile/inactive prostate is also a typical finding in juvenile dogs (as indicated by the terminology for this finding).

The evaluations by Charles et al. (1996a, 1996b) also support that findings from the rat and dog subchronic studies on 2,4-D acid are generally consistent with findings with the salts and esters of 2,4-D (when doses are expressed as acid equivalents), and that there are no unique toxicities, endocrine-mediated or otherwise, associated with these forms.

Epidemiological studies

Three limited epidemiological studies evaluating potential associations of 2,4-D and changes in human sperm parameters (Lerda & Rizzi 1991; Swan et al. 2003) or hormonal biomarkers (Garry et al. 2001) were identified. The Lerda and

Rizzi (1991) study is considered too limited in scope and relevant details, and is not considered to provide reliable evidence of male reproductive toxicity or endocrine disruption resulting from occupational exposure to 2,4-D. Swan et al. (2003) is considered too limited, due to the low numbers of control and case subjects with urinary 2,4-D levels above the LOD, to be considered in the WoE as evidence for presence or absence of an association. Garry et al. (2001) found no correlation between FSH, free testosterone, or total testosterone concentrations with 2,4 D urinary levels at the time of maximum 2,4-D usage. LH levels were reported to show a correlation, but the authors indicated the limited sample size warrants caution in drawing any conclusions from this study. It should also be noted that the animal studies showed no findings congruent with altered LH levels, such as increased Leydig cell tumors.

Based on a weight of evidence evaluation of the available data, including the absence of evidence for potential androgen pathway interactions in the Marty et al. (2010) Tier 2 EDSP equivalent EOGRT study, the lack of evidence for potential androgen pathway interactions predicted by the other key mammalian toxicity studies at doses below the TSRC, the absence of androgen-pathway-related responses in the FSTRA, the absence of adverse effects in the quail dietary reproductive toxicity study, and the negative Tier 1 EDSP *in vitro* AR binding, the negative ToxCast™ AR binding and AR transactivation and other high-quality published *in vitro* screening data and *in vivo* studies relevant to the androgen pathway, it is concluded that 2,4-D does not show evidence for direct interaction with the androgen pathway (either androgenic activity or anti-androgenic activity) at exposure levels relevant for human or ecological risk assessment.

WoE evaluation for the steraidogenic pathway or HPG axis interactions

Three EDSP Tier 1 screening assays relevant to steroidogenesis are available for 2,4-D, including the steroidogenesis (LeBaron et al. 2011c) and aromatase (Coady & Sosinski 2011) *in vitro* assays (published in Coady et al. 2014) and the FSTRA (Marino et al. 2010, published in Coady et al. 2013). The Marty et al. (2010) EOGRT study is an EDSP Tier 2 equivalent mammalian assay that also provides data possibly relevant to assessing whether 2,4-D interacts with the steroidogenic pathway, as do the key regulatory toxicity studies. The *in vivo* Tier 1 assay and key toxicity study results that may indicate potential interaction with the steroidogenesis pathway overlap substantially with those relevant to assessing potential interactions with the estrogen and androgen hormonal pathways, because the steroidogenic pathway is critical for the production of both estrogens and androgens. Positive effects in the *in vitro* steroidogenesis or aromatase assays provide supporting evidence that *in vivo* findings are influenced by interactions with these pathways, but negative findings do not rule out potential interactions. Endpoints and findings potentially relevant for addressing interactions of 2,4-D with the steroidogenic pathway are found in Tables 18-20.

Also relevant are selected parameters from the avian reproductive toxicity study, and results from published *in vitro* and *in vivo* studies. It should be noted that limited high quality published data are available investigating this potential mechanism. Supplementary information from the ToxCast™ aromatase assay is noted; no relevant aromatase or steroidogenesis assays were found in the peer reviewed literature.

Evaluation of potential interactions with the HPG axis is based on studies in intact mammals; relevant endpoints are not evaluated in other studies. Data are not separately tabulated to evaluate this interaction, as the gonadal findings are adequately captured by the evaluation for steroidogenesis interactions. Pituitary findings are discussed in the text.

As mentioned previously, the WoE for each pathway depends on multiple other factors. Changes in steroidogenic activity may be an underlying mechanism for changes in parameters that are also affected by androgen and/or estrogen pathways. The reader should be mindful that interaction with steroidogenesis pathway is also difficult to distinguish from findings secondary to systemic toxicity, because steroidogenesis depends on a complex and interrelated system of hormonal synthesis and feedback which may be influenced markedly by factors such as a decrease in the cholesterol starting material (from poor nutrition or disruption of synthesis or metabolism in the liver) or changes in membrane transport of steroid precursors, changes in mitochondrial function or other effects mediated by oxidative stress or through other membrane toxicity. Non-specific stress due to excessive toxicity or non-compound-related factors such as immobilization may also affect the steroidogenic process (Orr et al. 1994; Orr & Mann 1990). Adrenal weight and specific histopathological changes may point toward a steroidogenic or stress-related mechanism, but it should be noted there are multiple sites for steroidogenesis in the intact organism.

Note there are no exposure-related or equivocal findings that suggest altered steroidogenesis for 2,4-D in mammalian studies below the TSRC, and no specific evidence of modulation of steroidogenesis in the FSTRA.

In vitro *studies*

In the Tier 1 EDSP steroidogenesis assay (LeBaron et al. 2011c, published in Coady et al. 2014), there was a slight increase in estradiol in all runs at the high concentration only (100 µM). The 1.2-fold increase was below the 1.5-fold response threshold established in the EDSP steroidogenesis validation assays as a positive response (Hecker et al. 2008), and therefore this finding is not considered biologically meaningful.

The Tier 1 EDSP aromatase assay (Coady & Sosinski 2011, published in Coady et al. 2014) was negative. The ToxCast™ aromatase assay developed under the auspices of the US EPA also showed no evidence of aromatase inhibition.

The published *in vitro* literature lacked high quality studies evaluating potential effects on steroidogenesis or aromatase.

Based on the above data, it is unlikely that 2,4-D affects steroidogenesis *in vitro*.

Table 18. Results for 2,4-D from EDSP tier 1 *in vitro* toxicology and ecotoxicology assays relevant to potential interaction with steroidogenesis.

Study*	Steroidogenesis	Aromatase inhibition	Fish fecundity	Fish fertility	Fish nuptial tubercles (M)	Fish nuptial tubercles (F)	Fish gonad somatic index (M)	Fish gonad somatic index (F)	Fish gonad histopathology (M)	Fish gonad histopathology (F)	VTG (M)	VTG (F)
Steroidogenesis	[redacted]	–	–	–	–	–	–	–	–	–	–	–
Aromatase	–	[redacted]	–	–	–	–	–	–	–	–	–	–
FSTRA	–	–	L↓ (3)‡	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

*Steroidogenesis: OPPTS 890.1550 Steroidogenesis (LeBaron et al. 2010); Aromatase: OPPTS 890.1200 Aromatase (Coady & Sosinski 2011); FSTRA: OPPTS 890.1350 Fish short-term reproduction assay (Marino et al. 2010).

†LeBaron et al. 2010: Steroidogenesis assay showed slight increased estradiol in all runs at the high concentration. The fold increase was below that established in the EDSP validation assays as a positive response (Hecker et al. 2006), and is therefore not considered biologically meaningful.

‡Marino et al. 2010: Decreased fecundity observed only at high (limit dose) concentration; non-specific finding.

All studies

NNo evidence of interaction

LEquivocal evidence of potential interactionat the limit dose only

-Endpoint not evaluated

↓Decreased relative to control

MMale

FFemale

Endpoint scores in parentheses based on Borgert et al. 2011 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

Ecotoxicological studies

The FSTRA (Marino et al. 2010; Coady et al. 2013) showed no findings likely to be associated with increased or decreased testosterone or estradiol. Secondary sex characteristics, fertility, gonadal somatic index (GSI), gonadal histopathology, and VTG levels were not affected in this study in males or females. There was only the single finding of decreased fecundity, at the high concentration only, which could potentially be associated with altered steroidogenesis. As previously discussed, this is a non-specific finding and is considered likely confounded by stress or uncharacterized systemic toxicity. As the steroidogenesis pathway is generally considered well conserved among vertebrate species, the absence of effects on this pathway in the EOGRT study also supports the concept that the non-specific decreased fecundity in the high concentration group of the FSTRA is likely due to stress or uncharacterized systemic toxicity and not to a potential interaction with steroidogenesis. While there are differences in exposure route (oral compared to via the gills), 2,4-D is completely absorbed via the oral route in rats and would be expected to be similarly readily absorbed through the gills. Further, both fish and rats would be exposed to parent 2,4-D as 2,4-D would avoid first-pass liver metabolism in fish due to the route of entry, and 2,4-D is not highly metabolized in rats.

A high dose reproductive toxicity study in quail (Mitchell et al. 2000) does not provide evidence of any effects that could be associated with increased or decreased testosterone or estradiol.

In published studies, a steroidogenesis assay using alligator eggs directly exposed to 2,4-D has been performed, which showed no effect of 2,4-D on steroidogenesis (Crain et al. 1997). This study used unconventional methodology but is of interest because it expands the range of species tested. Estradiol was used as a positive control, resulting in development of ovaries in embryos incubated at male-producing temperatures that was also associated with increased gonadal–adrenal mesonephros complex aromatase activity.

Mammalian studies

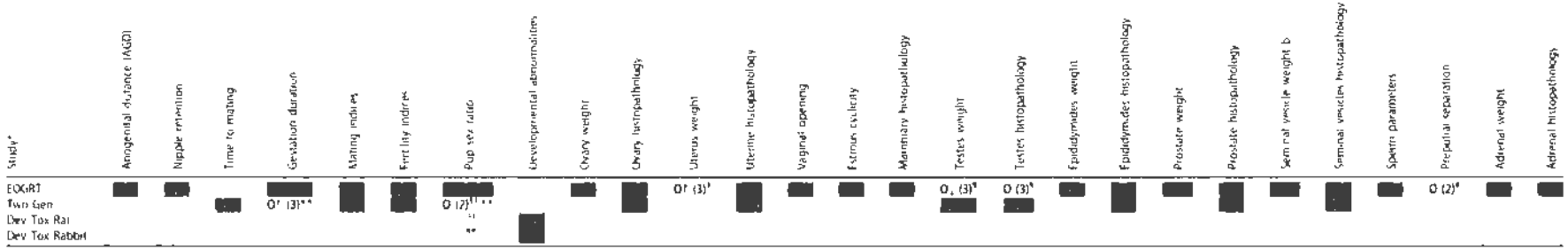
There were no effects in the EOGRT study (Marty et al. 2010, published in Marty et al. 2013) suggesting increased or decreased estradiol:

- There were no exposure-related effects on developmental landmarks, including AGD (measured in all F1 pups), or age at vaginal opening (measured in all F1 Set 1–3 females);
- There were no effects on estrous cycle length or estrous cycle pattern (evaluated in all P1 main study and satellite females and all Set 3 F1 females), including a lack of persistent estrus, at any dose level.
- There were no exposure-related effects on reproductive indices, including mating, fertility, time to mating, gestation length, pre- and post-implantation loss and corpora lutea number (satellite group).
- There were no signs of dystocia in 2,4-D-exposed P1 dams.
- Litter size and pup survival were not affected by 2,4-D in this study.
- There were no biologically significant exposure-related effects on reproductive organ weights at any dose of 2,4-D: no statistically significant changes in uterine weight, or high dose uterine weights outside of the laboratory HCD range; and
- 2,4-D did not alter reproductive organ or mammary gland histopathology, including male mammary gland histopathology and ovarian follicle counts in F1 females with the longest duration of exposure, and exposure during critical windows of development.

Overall, the data from this detailed EOGRT in CD®SD rats do not support any 2,4-D-mediated effect on estradiol synthesis, even at doses exceeding the TSRC for 2,4-D in rats.

In the key regulatory toxicity studies, including a two-generation reproductive toxicity study (Rodwell & Brown 1985), a developmental toxicity study (Rodwell 1983), and subchronic

Table 19. Results for 2,4-D from regulatory reproductive and developmental toxicity studies relevant to potential interaction with steroidogenesis.



*EOGRT: EOGRT study (Marty et al. 2010); Two Gen: OPP 83-4 Two-generation reproductive toxicity study (Rodwell & Brown 1965); Dev Tox Rat: OPP 83-3 developmental toxicity study in rats (Rodwell 1983); Dev Tox Rabbit: OPP 83-3 developmental toxicity study in rabbit (Hoberman 1990).

†With coagulating gland.

‡Marty et al. 2010: High dose uterine weight increases not considered exposure related (non-statistically significant change in cycling females; no correlating histopathology except normal estrous cycle related changes; within HCD).

¶Decreased testes weight F1 pups at weaning, attributable to decreased body weight, not apparent in F1 adults.

§Testicular atrophy in 2 high-dose P1 males, considered spontaneous, not present in F1 dosed animals.

||Seminal vesical and prostate weights decreased compared to control at ~300 ppm in P1; not statistically significant, within HCD and controls below HCD; conclusion no effect.

#Increased age at preputial separation; considered associated with decreased growth and not endocrine related.

**Rodwell and Brown 1965: The length of gestation was statistically significantly prolonged (by 1 day) in the production of the F1b pups at ≥80 mg/kg/day, compared with controls; likely attributable to the very excessive dose of 2,4-D to the dams during production of the F1b litters; high pup mortality was seen at this excessive dose (~100 mg/kg/day).

††80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (increased M pups), compared with the controls. This finding was not repeated in the F1b pups (at a higher dose) and is considered unlikely to be exposure related because of the lack of consistency across generations or with other 2,4-D studies.

‡‡Evaluated but not sensitive; as offspring sex was determined prior to initiation of dosing.

¶¶Change in fetal sex ratio present; not attributed to test article as offspring sex determined prior to initiation of dosing; no selective loss of females

All studies:

NNo evidence of potential interaction

OFinding over the threshold for renal saturation

–Endpoint not evaluated

↑Increased relative to control

↓Decreased relative to control

Endpoint scores in parentheses based on Borgert et al. 2011 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

Table 20. Results for 2,4-D from regulatory subchronic (SC) and chronic (C) toxicity studies relevant to potential interaction with steroidogenesis.

Study*	Ovary weight	Ovary histopathology	Uterine histopathology	Vaginal histopathology	Mammary histopathology	Testes weight	Testes histopathology	Epididymides weight	Epididymides histopathology	Prostate histopathology	Seminal vesicles histopathology	Adrenal weight	Adrenal histopathology
Rat SC (1)	O↑ (2)†				–	O↓ (2)‡	O (2)‡			–	–	O(2)¶M↑F↓	O (2)¶ MF
Rat SC (2)	–			–	N(2)*	O↓ (2)‡		–		N(2)	N(2)	–	
Rat SC (3)	–	–	–	–	–			–	–	–	–	–	–
Rat C	O↓ (3)§	N(2)	N(2)	N(3)	N(3)	O↓ (2)‖	O (2)‡	–	N(2)	N(2)	N(2)	O↓ (2)**	
Mouse SC	N(2)*	N(2)	N(2)	–	–			–	N(2)	–	–	O↑(2)††	
Mouse C (1 F)	–	N(3	N(2)	N(3)	N(3)	–	–	–	–	–	–	–	
Mouse C (2 M)	–	–	–	–	–	O↑ (2)‡‡	N(2)	–		N(2)	N(2)	–	
Dog SC (1)		N(2)		–	–	O↓ (3)¶¶	O (3)§§	O↓(3)‖‖		–	–	–	
Dog SC (2)		N(2)				O↓ (3)##		–			–		
Dog C		–						–			–		

*Rat SC (1): OPP 82-1 13-week rat subchronic toxicity study (Schulze 1991a); Rat SC (2): 13 week rat subchronic toxicity study 82-1 (Gorzinski et al. 1981a); Rat SC (3): 13 week rat subchronic toxicity study (Gorzinski et al. 1981b); Rat C: OPP 83-5 Two year rat chronic/oncogenicity study (Jeffries et al. 1995); Mouse SC: OPP 82-1 13 week mouse subchronic toxicity study (Schulze 1991b); Mouse C(1 F): OPP 83-2 mouse oncogenicity study (females) (Stott 1995a); Mouse C(2 M): OPP 83-2 mouse oncogenicity study (males) (Stott 1995b); Dog SC(1): OPP 82-1: 13-week dog subchronic toxicity study (Schulze 1990); Dog SC(2): OPP 82-1: 13-week dog subchronic toxicity study (Dalgard 1993a); Dog C: OPP 83-1: dog chronic toxicity study (Dalgard 1993b).

†Schulze 1991a Increased ovary weight at 300 mg/kg/day (far exceeded TSRC and systemically toxic), no histopathological correlate.

‡Schulze 1991a and Gorzinski et al. 1981a Testes weight: decrease at ≥100 mg/kg/day; atrophy at 300 mg/kg/day (far exceeded TSRC and systemically toxic).

¶Schulze 1991a Adrenal: weight changes and histopathology (hypertrophy *zona glomerulosa*) at 300 mg/kg/day (far exceeded TSRC and systemically toxic); *zona glomerulosa* does not respond to HPA axis.

§Jeffries et al. 1995 Ovary: weight decrease at high dose attributable to body weight decrease; no correlating histopathological changes.

‖Testes: absolute and relative testes weights decreased at 150 mg/kg/day at both interim and terminal sacrifices; no corresponding histopathological findings at terminal sacrifice.

#Testes: 2/10 with atrophy at intermediate (1 year) sacrifice in high-dose group; no exposure related findings at 2 yr sacrifice.

**Adrenal: weights in females decreased at 75 and 150 mg/kg/day at 2 yr sac.; may be attributable to body weight decreases.

††Schulze 1991b Adrenal weights increased; no dose response; considered unlikely to be exposure related.

‡‡Stott 1995b Relative testes weight increased but not absolute weight at 24 month only; attributed to body weight decrease.

¶¶Schulze 1990 Testes weight decreased at 10 mg/kg/day (dose exceeded MTD).

§§Hypospermia in 2 of the 10 mg/kg/day dose group surviving dogs; equivocal relationship to exposure (dose exceeded MTD).

‖‖Weight taken with testes weight; see comments on testes; no histopathological correlates in epididymides.

##Dalgard 1993a Testes weight decreased at 7.5 mg/kg/day; no correlating histopathological changes

All studies:

NNo evidence of potential interaction

OFinding *over* the threshold for renal saturation

–Endpoint not evaluated

↓Decreased relative to control

↑Increased relative to control

Endpoint scores in parentheses based on Borgert et al. 2014 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

Table 21. Data from 2,4-D tier 1 EDSP and regulatory toxicity studies relevant to potential interactions with the HPT axis.

Study*	Tadpole developmental stage (day 7 and 21)	Tadpole normalized hind limb length (day 7 and 21)	Asynchronous development (day 7 and 21)	Thyroid weight	Thyroid histopathology	Thyroid hormone (T4)	Thyroid hormone (T3)	Thyroid hormone (TSH)	Pituitary weight	Pituitary histopathology
AMA				–		–	–	–	–	–
EOGRTS	–	–	–	↑(2)	O (1)†	O↓ (2)¶	O↓(2)¶	O↓(2)§	O↓ (3)‖	
Rat SC (1)	–	–	–	O↑ (2)#	O (1)**	O↓ (2)	O↓(2)	–	O M ↑; F↓ (3)††	
Rat SC (2)	–	–	–	–		O↓ (2)	–	–	–	
Rat SC (3)	–	–	–	–	–	O↓ (2)	–	–	–	–
Rat C	–	–	–	O↑ (2)‡‡	O (1)¶¶	O↓ (2)§§	–	–	–	
Mouse SC	–	–	–	↑(2)	↑(1)	O↓ (2)‖‖	–	–	N(3)	
Mouse C (1 F)	–	–	–	–		–	–	–	–	
Mouse C (2 M)	–	–	–	–		–	–	–	–	–
Dog SC (1)	–	–	–					–	–	
Dog SC (2)	–	–	–	O↓(2)##		–	–	–		
Dog C	–	–	–			–	–	–		

*AMA: amphibian metamorphosis assay (Coady et al. 2013); EOGRTS: F1-extended one generation dietary toxicity study (Marty et al. 2010); Rat SC (1): OPP 82-1 13-week rat subchronic toxicity study (Schulze 1991a); Rat SC (2): OPP 82-1 13-week rat subchronic toxicity study (Gorzinski et al. 1981a); Rat SC 3: 13-week rat subchronic study (non-guideline) (Gorzinski et al. 1981b); Rat C: OPP 83-5 two year rat chronic/oncogenicity study (Jeffries et al. 1995); Mouse SC: OPP 82-1 13 week mouse subchronic toxicity study (Schulze 1991b); Mouse C(1 F): OPP 83-2 mouse oncogenicity study (females) (Stott 1995a); Mouse C(2 M): OPP 83-2 mouse oncogenicity study (males) (Stott 1995b); Dog SC (1): OPP 82-1: 13-week dog subchronic toxicity study (Schulze 1990); Dog SC (2): OPP 82-1: 13-week dog subchronic toxicity study (Dalgard 1993a) and Dog C. OPP 83-1: dog chronic toxicity study (Dalgard 1993b)

†Marty et al. 2010 Thyroid weight change not considered exposure related; no dose response, no consistency in direction or biologically significant magnitude of change and no correlating histopathology.

‡Thyroid histopathology slight adaptive change (depleted colloid) in GD 17 females at 600 ppm.

¶Non-statistically significant decrease T3 and T4 at 600 ppm; considered likely exposure related in GD 17 females because of pattern of findings.

§Non-statistically significant increase TSH at 600 ppm; considered likely exposure related in GD 17 females because of pattern of findings.

‖Decreased pituitary weight in set 3 males only; no correlating histopathology; considered unlikely to be exposure related.

#Schulze 1991a: thyroid weight: Males had higher absolute thyroid/parathyroid weights at 300 mg/kg/day and relative weights at 100 and 300 mg/kg/day Females at 300 mg/kg/day had higher relative thyroid/parathyroid weights.

**Follicular cell hypertrophy (adaptive change) in females at high and systemically toxic 300 mg/kg/day dose exceeding MTD and renal clearance saturation threshold.

††Females had lower absolute and relative pituitary weights at 300 mg/kg/day. Males had higher relative pituitary weights at 300 mg/kg/day.

‡‡Jeffries et al. 1995 Thyroid weights (absolute and relative) increased in males at 150 mg/kg/day, and in females at 75 and 150 mg/kg/day at both sacrifices.

¶¶Decreased colloid (adaptive change) in thyroid of females at 1 year at 150 mg/kg/day; parafollicular hyperplasia non-statistically significantly increased at 150 mg/kg/day.

§§Decreased T4 at 150 mg/kg/day at 1 year interval.

‖ Schulze 1991b Non statistically significant decreased T4 at 100 and 150 mg/kg/day.

##Dalgard 1993a Relative but not absolute weight increase at high dose (7.5 mg/kg/day) attributed to body weight loss

All studies

NNo evidence of potential interaction

OFinding *over* the threshold for renal saturation

–Endpoint not evaluated

↑Increased relative to control

↓Decreased relative to control

MMale

FFemale

Endpoint scores in parentheses based on Borgert et al. 2014 (modified in some cases based on context or strength of response)

(1) specific and sensitive to the hypothesis

(2) potentially sensitive for the hypothesis; stronger if correlated with Rank 1 data

(3) relevant, but useful only if corroborating Rank (1) or (2) endpoints

studies, few endpoints were observed suggesting increased or decreased estradiol, even at dose levels causing significant systemic toxicity.

Ovary weights were increased at 300 mg/kg/day in the Schulze (1991a) subchronic rat study; there were no histopathological findings in the ovaries correlating with this change. No other changes suggesting increased estradiol were seen in this study, nor were any findings in females suggesting decreased estradiol. In contrast, the Jeffries et al. (1995) chronic rat study showed decreased ovarian weights at the high dose (which caused marked systemic toxicity and weight loss), again without histopathological correlates. The Jeffries et al. study also showed a decreased incidence of pituitary tumors of the *pars distalis*, which is an estrogen-sensitive tumor (Dinse et al. 2010), and a decreased incidence of mammary gland hyperplasia in 2,4-D exposed rats, also at the severely toxic high dose. The extent of weight loss at this dose confounds any attribution of these findings to decreased estradiol. The mouse subchronic (Schulze 1991b) and chronic studies (Stott 1995a) showed no effects on uterine or ovarian histopathology. The dog subchronic (Schulze 1990; Dalgard 1993a) and chronic (Dalgard 1993b) studies similarly showed no effects on uterine, vaginal or mammary gland histopathology.

There were no effects in the Marty et al. (2010) EOGRT study considered indicative of either increased or decreased testosterone:

- Preputial separation was slightly delayed in the F1 males at 800 ppm, which was attributed to decreased growth during lactation and post-weaning, as discussed previously;
- There were no exposure-related effects on developmental landmarks, including AGD (measured in all F1 pups),

nipple retention (measured in non-culled F1 pups in all dose groups);
- There were no effects on sperm counts, motility or morphology;
  There were no exposure-related effects on reproductive indices, including mating, fertility and time to mating;
- There were no exposure-related effects on reproductive organ or accessory sex tissue weights at any dose of 2,4-D; and
- 2,4-D did not alter reproductive organ or accessory sex tissue histopathology.

There were comparatively a few relevant parameters evaluated in the Rodwell and Brown (1985) two-generation study. Mating and fertility were not affected, nor were testicular weights and histopathology, or histopathology of the epididymides, seminal vesicles (weanlings only) and prostate (weanlings only).

There were no exposure-related effects on testicular weight or histopathology in the Marty et al. (2010) EOGRT study attributable to interactions with steroidogenesis, or in the Rodwell and Brown (1985) two-generation reproductive toxicity study. A single finding in the Rodwell and Brown (1985) two generation study potentially showing a steroidogenesis interaction was that the 80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (109 males and 71 females), compared with the controls. As discussed previously, this finding was not repeated in the F1b pups at a higher dose and is considered unlikely to be exposure-related.

The Schulze (1991a) subchronic rat study showed flaccid testes and testicular atrophy at the very high and systemically toxic dose of 300 mg/kg/day (that exceeded both the TSRC and a classic MTD). The high-dose testes findings could hypothetically be associated with changes in steroidogenesis (excess estradiol and/or decreased testosterone) secondary to systemic toxicity; they could as readily reflect changes to the HPG axis caused by the severe systemic toxicity at these excessive doses, or reflect direct target organ toxicity not mediated by an endocrine interaction, such as that related to stress. Given that there is no substantive evidence of increased estradiol or decreased testosterone in the 2,4-D mammalian studies as a whole, the stress or systemic toxicity seem the most likely hypotheses. Decreased testes weight was also seen at the systemically toxic high dose in the chronic rat study (Jeffries et al. 1995). Testicular atrophy was noted in 2/10 animals at the interim sacrifice in this study; however, no exposure-related testicular lesions were evident at the terminal (two-year) sacrifice.

Mice in a subchronic study (Schulze 1991b) showed no effects on male reproductive tissues. Increased testicular weight was seen at the high dose in a chronic mouse study (Stott 1995b); there was no histopathological correlate to the testicular weight finding.

Dogs in 2,4-D subchronic toxicity studies (Schulze 1990; Dalgard 1993a) showed decreased testes weights at systemically toxic doses; as discussed previously, this finding is not considered likely to be endocrine-mediated, but rather to reflect a combination of the immature age of the test animals and, possibly, delayed development due to systemic toxicity. These findings were at doses exceeding the TSRC in dogs. There were no testicular or prostate effects in the chronic dog study (Dalgard 1993b). As discussed previously, dogs are not relevant for human health risk assessment; however, as a susceptible species, the dog may predict potential effects on other species deficient in the OAT-1 transporter. Therefore, data from the dog studies are included in the WoE evaluation. It should be noted particularly that all potentially endocrine-related effects in the dogs were seen at dose levels that also caused other marked systemic toxicity. An EPA Science Advisory Panel has agreed with the EPA position that responses observed in endocrine disruption assays in the presence of overt toxicity are "not useful for interpretation of whether a compound has an endocrine effect" (US EPA 2013). Thus, there does not appear to be any particular susceptibility to potentially endocrine-related effects.

There were no effects on adrenal weight or histopathology in the EOGRT study (Marty et al. 2010). The Schulze (1991a) rat subchronic study showed adrenal weight changes and histopathology (hypertrophy of the *zona glomerulosa*) at 300 mg/kg/day (far above the TSRC and systemically excessively toxic). The adrenal *zona glomerulosa* is responsible for production of mineralocorticoids such as aldosterone and does not respond to changes in the hypothalamic–pituitary–adrenal (HPA) axis. Consequently, this finding is not considered evidence of an interaction with the steroidogenesis pathway. Adrenal weight changes (in opposite directions) were also seen in the mouse subchronic study (Schulze 1991b) and in the chronic rat study (Jeffries et al. 1995). No exposure-related histopathology was found in the adrenals in these studies.

In conclusion, 2,4-D does not show robust evidence of interaction with the steroidogenesis pathway(s) at environmentally relevant exposure levels. Mammalian studies, including a comprehensive EDSP Tier 2 equivalent EOGRT study, fail to show coherent evidence of alterations in estradiol synthesis or testosterone synthesis at doses below the TSRC. Even at high doses, findings are limited and may reflect direct target organ toxicity without an endocrine-mediated mechanism, e.g. effects associated with excessive systemic toxicity. There are no robust effects in the FSTRA indicating altered steroidogenesis, and a quail reproductive toxicity study showed no findings suggesting altered steroidogenesis. The *in vitro* EDSP steroidogenesis and aromatase assays were negative, as was the ToxCast™ aromatase assay.

Evaluation of potential interaction with the HPG axis

The majority of parameters potentially under control of the HPG axis were unaffected in the Marty et al. (2010) EOGRT study as discussed above. Absolute and relative (fixed) pituitary gland weights were significantly decreased by 9 and 8%, respectively, in the high dose males in one set of F1 adult animals. The magnitude of the differences from pituitary weights in control animals, however, was extremely slight and the absolute and relative pituitary weights in the 800 ppm males were within the historical control range. No exposure-related pathological changes were seen in these tissues. Additionally, toxicologically significant alterations in pituitary function would be expected to alter numerous other

study endpoints, including reproductive and accessory sex gland weights and sperm parameters. These endpoints were not affected by 2,4-D exposure in this study. It is concluded that this study shows no robust evidence of an HPG axis interaction.

There is limited and inconsistent evidence of an HPG axis interaction in the other key toxicity studies of 2,4 D. As previously discussed, in the Rodwell and Brown (1985) reproductive toxicity study, there was an altered pup sex ratio in the F1a litters; this finding as noted above is considered unlikely to be exposure related. Pituitary weights were decreased in females in the Schulze (1991a) rat subchronic toxicity study at the high and systemically toxic dose of 300 mg/kg/day (exceeding an MTD and far exceeding the TSRC) but were increased in males; there were no histopathological correlates in either sex and the exposure relationship is considered equivocal. A decrease in tumors in the *pars distalis* of the pituitary was seen in the Jeffries et al. (1995) rat chronic toxicity study; the decreased incidence of this estrogen responsive tumor is attributed to the marked weight loss at the excessively toxic high dose.

WoE evaluation for potential effects of 2,4-D on the HPT axis

The HPT axis is an integrated system involving various positive and negative feedback systems to control production of thyroid hormones. To characterize these feedback systems, the use of an intact, *in vivo* model is required. For evaluation of potential HPT axis interaction, it is useful to characterize whether (if there are effects) the compound is acting as a thyroid agonist or antagonist. Agonists show increases in circulating thyroid hormone levels (e.g. T4) but may or may not impact other thyroid-related parameters. Strong agonists or antagonists would also be anticipated to influence clinical and behavioral observations (Fliers et al. 2006; Helmreich & Tylee 2011). Stress may also impact the HPT axis (*ibid*); so, caution should be taken when evaluating findings in the presence of other systemic toxicity or when manipulating animals for thyroid hormone collection. Antagonists, which are much more frequently identified because of the multiple mechanisms that can lead to decreased circulating thyroid hormones (and also due to the particular susceptibility of rats to this effect) usually show a pattern of feedback-related changes including T4 decreases and/or T3 decreases and feedback-mediated TSH increases, which typically result in follicular cell hypertrophy of the thyroid gland, thyroid gland weight increases and follicular cell hyperplasia (Marty et al. 2001). Potent anti-thyroid agents may also result in developmental neurotoxicity if exposures are pre- or peri-natal (see for example Goldey et al. 1995; Shibutani et al. 2009; Bernal 2012).

One *in vivo* Tier 1 EDSP screening assay available for 2,4-D, the AMA (Coady & Sosinski 2011), is considered important to assess potential interactions of 2,4-D with the HPT axis, because amphibian development is very much under thyroid control. The Marty et al. (2010) Tier 2 equivalent EOGRT dietary toxicity study provides detailed information on potential thyroid toxicity of 2,4-D, and also provides information on potential developmental neurotoxicity findings, including brain morphometry and neuropathology, myelin deposition, hypothalamus histopathology and auditory startle, that might be affected if thyroid function was impaired. Additional information is available from the subchronic toxicity studies of 2,4 D. These assays and the endpoints and results relevant to potential interaction with the HPT axis are summarized in Table 21.

This WoE analysis for the HPT axis is based on the results from EDSP Tier 1 screening and other available studies summarized above.

Note there were no findings for 2,4 D considered exposure related and relevant for assessment of the HPT axis at doses below the TSRC, and findings at higher doses were primarily considered adaptive and non-adverse.

There were no exposure-related effects in the AMA (Coady et al. 2010), which is designed to specifically identify potential thyroid effects. This study tested 2,4-D up to a concentration approximating the limit dose (100 mg/L).

In the EOGRT study (Marty et al. 2010), there were many slight changes in hormone levels and/or thyroid organ weight that did not reach statistical significance or show a pattern of findings typically associated with altered thyroid function, and that were not accompanied by thyroid histopathological changes. As a result, these findings are considered non-adverse and unlikely to be exposure related.

Only one lifestage in the Marty et al. (2010) EOGRT dietary toxicity study shows a pattern of effects, but only at the highest dose that exceeds the TSRC, which appears possibly exposure-related. In GD 17 females, there was a dose-related *pattern* of non-statistically significantly decreased T3 and T4, increased TSH and slight evidence of thyroid histopathology (reduced colloid) at the 600-ppm dose level suggesting an adaptive exposure-related effect. There is a mechanistic basis for this finding, in that high dose 2,4-D has been shown to displace thyroxine from serum binding protein in the rat (Van den Berg et al. 1991). This displacement could lead to easier excretion or hepatic sequestration of the free hormone. Further, thyroids in dams are stressed during gestation by the need to supply thyroid hormone to the developing fetuses, making the dams relatively hypothyroid and vulnerable to such an effect. In addition, because of increased dosed feed consumption during gestation and the lack of dietary concentration adjustment during this critical stage, dams were receiving 2,4-D at a significantly higher internal dose than animals did at most other time points in the study. The lack of adversity is demonstrated by: the lack of replication in dams at lactation day 22, showing reversibility; the slight severity of the findings; and the lack of adverse findings that might be associated with decreased thyroid function in the F1 pups. For example, there were no findings in the developmental neurotoxicity (DNT) component of the EOGRT study (Marty et al. 2010) consistent with thyroid hormone modulation. No exposure related effects on auditory startle, brain morphometric or myelin deposition changes were seen, demonstrating the lack of an adverse thyroid hormone deficiency during fetal and pup development.

The Schulze (1991a) subchronic rat study showed stronger effects on thyroid hormone economy at doses

≥100 mg/kg/day. Gorzinski et al. (1981a, 1981b) showed decreased T4 at 100 mg/kg/day. Schulze (1991a) also identified exposure-related effects on thyroid histopathology in female rats at a dose exceeding an MTD (300 mg/kg/day); even then, the histopathological effect was limited to follicular cell hypertrophy, which may be regarded as adaptive. Mice in the subchronic study (Schulze 1991b) showed decreased T4 at the high dose but no changes in thyroid weight or histopathology; mice in the chronic studies (Stott 1995a, 1995b) showed no effect on thyroid histopathology in either females or males.

In contrast to rodent study results, the subchronic dog study by Schulze (1990) tested dogs to doses above the MTD and saw no consistent effects on the thyroid. (Relative thyroid weight was increased at the high dose, but absolute weight was not affected; this finding is attributed to body weight loss at the high dose.) There were no effects in this study on thyroid hormone measurements (T3 and T4), nor on thyroid histopathology. The other subchronic dog study (Dalgard 1993a) and the chronic dog study (Dalgard 1993b) showed no effects on thyroid weights or histopathology, despite testing systemically toxic doses exceeding the dog TSRC. The dog studies provide data supporting that rodent species (especially rats) are particularly vulnerable to changes in thyroid hormone economy.

Charles et al. (1996a) presented data from rat subchronic toxicity studies conducted with 2,4-D DMA, or 2,4-D 2-EHE. Endocrine endpoints relevant to the thyroid pathway included: thyroid hormones (T3 and T4); thyroid (and parathyroids) organ weights and histopathological evaluations. Decreased T4 and/or T3 was observed at dose levels of >100 mg/kg/day, with T4 appearing somewhat more sensitive than T3 and females more sensitive than males. Correlating with these findings were increases in relative thyroid weights (primarily at 300 mg/kg/day); however, no correlating histopathological evidence of thyroid follicular cell hypertrophy or hyperplasia was evident at any dose. Therefore, these changes are considered slight in severity and non-adverse.

Charles et al. (1996b) also described data from dog subchronic toxicity studies conducted with 2,4-D DMA, or 2,4-D 2-EHE. Endocrine endpoints evaluated in these studies relevant to the HPT axis included pituitary, thyroid (and parathyroid) organ weights and histopathological evaluations. Thyroid hormone analyzes were not performed in these studies, which is a weakness for evaluating potential thyroid effects. In contrast to the rat, there were no findings in dogs supporting an interaction with the HPT axis, even though the high doses in these studies were markedly systemically toxic.

Discussion of mechanism of high dose-specific effect on HPT axis in rodents

Plasma protein binding appears to protect circulating thyroid hormone from metabolism and clearance by the liver. Thus, if 2,4-D is a relatively weak competitor with thyroid hormone for binding sites or transport/carrier protein, high exposures to 2,4-D would be anticipated to result in increased free thyroid hormone which would be subject to enhanced sequestration and/or excretion by the liver. This would not be an outcome anticipated at low exposure levels, however, which is consistent with the thyroid-related findings being limited to high-dose 2,4-D exposure in the rodent subchronic and chronic toxicity studies summarized above.

The Florsheim et al. (1963) study supports the idea that high doses of 2,4-D in the rat may modulate thyroid hormone levels. These data, in conjunction with the Van den Berg et al. (1991) study, support that the likely primary mechanism is 2,4-D competition for the thyroxine serum binding sites, particularly transthyretin. It should be noted that even in the chronic rat toxicity study of 2,4-D (Jeffries et al. 1995) there was no frank progression to thyroid follicular cell hyperplasia or neoplasia, suggesting changes in circulating thyroid hormone levels were sufficiently mild to not elicit biologically adaptive and sustained elevations of TSH. Therefore, these findings provide a possible mechanistic explanation for decreases in circulating thyroid concentrations in rats selective to high dosages of 2,4-D, but do not provide evidence of a biologically significant adverse effect.

The rat is more likely to be susceptible to this mechanism than the human because the predominant rat thyroid hormone binding protein binds thyroxine less tightly than that of the human. In humans and other primates, thyroxine-binding globulin (TBG) is the principal protein that binds T4 (Dohler et al. 1979). It has a very high affinity for T4: only about 0.03% of the T4 in serum is in the free unbound form (Hill et al. 1989). Binding sharply reduces clearance of T4 from serum. Rats do not have TBG, and most T4 in rat serum is bound to albumin and transthyretin. The binding affinity of T4 for TBG is more than a 100 times greater than that of albumin or transthyretin (Hill et al. 1989), and the difference contributes to the higher rate of T4 clearance in rats. Further evidence that the rat is an overly sensitive species is the lack of thyroid findings in the subchronic and chronic 2,4-D dog studies, conducted at doses clearly exceeding the TSRC in that species.

Overall, there are no findings in the 2,4-D studies suggesting an adverse effect on the thyroid or clear evidence for an HPT axis interrelationship at doses below the TSRC. Adaptive changes were seen in pregnant dams during a susceptible life stage in the Marty et al. 2010 EOGRT study (also at a dose exceeding the TSRC); the mechanism for these high-dose specific findings has been characterized. No adverse effects were observed on offspring, either for thyroid parameters or in assessment of potential developmental neurotoxicity. The thyroid does not appear to provide a POD or driving effect for 2,4-D risk assessment because more sensitive indicators of toxicity are present, which occur within the linear TK range. It should be further noted that the serum protein binding to thyroxine in humans is considerably stronger than in rats (Jahnke et al. 2004), providing an extra margin safety for humans to any potential thyroid toxicity.

Interestingly, despite reports of good interspecies concordance for the HPT axis, the AMA (Coady et al. 2010) was negative up to the limit dose tested. This lack of concordance may relate to 2,4-D's postulated mechanism for thyroid hormone effect, which is displacing bound thyroxine from the binding proteins used for systemic transport, making the thyroxine more available for excretion or hepatic storage. It is

reasonable to hypothesize that the affinity of the binding protein to circulating thyroid hormone in frogs is different from the affinity of the binding protein in rat and that the difference in interspecies response might be attributable to that difference; no research directly addressing potential differences has been identified.

The WoE shows that, though there is weak evidence of 2,4-D potentially interacting with the HPT axis, this is very unlikely to result in any adverse effects at exposure below the relatively high doses characterizing the onset of the TSRC, even in rodents. No adverse effects on the thyroid, or adverse sequelae to the offspring, including effects on myelination or brain morphometric parameters, were identified in the Tier 2 equivalent EOGRT study up to the highest dose tested. The interaction has been studied across life stages and there is a high degree of confidence in this conclusion. Exposure to 2,4-D did not result in any thyroid-related effects in frogs tested up to the limit dose.

The Goldner et al. (2013) assertion of biological plausibility for a specific association of 2,4-D with hypothyroidism or thyroid disease in humans is very tenuous. Although the hypothyroid associations reported in Goldner et al. (2013) included positive associations with multiple herbicides and insecticides, the Stoker "paper" used by Goldner et al. to justify biological plausibility of their reported epidemiological findings specifically for 2,4 D is an abstract of an extremely high-dose study conducted in rats (100 and 200 mg/kg/day by oral gavage) which reported reductions in circulating thyroid hormone at both of the very high doses. This is consistent with findings in regulatory rodent toxicity studies of high dose decreases in T4 and adaptive changes in thyroid histopathology (limited to colloid depletion and in some cases hypertrophy, without evidence of hyperplasia or follicular cell tumors) at doses substantially exceeding the TSRC. Importantly, however, Stoker later published an abbreviated summary of these findings in a book chapter (Stoker & Zorrilla 2010) in which it was noted that the 2,4-D thyroid effects were not detected at the next lower dose of 30 mg/kg/day: "... and the herbicide 2,4-diphenoxyacetic acid (2,4-D) (sic), which induced renal toxicity at both 3 and 30 mg/kg and did not alter thyroid hormone (T4) or any of the other male pubertal endpoints until 100 mg/kg ....".

The Stoker and other related toxicity and biomonitoring data are thus not causally supportive of human thyroid disease, and in fact demonstrate an extremely low biological plausibility for any such outcome for the following reasons. First, it is well established that oral gavage doses of 100 mg/kg are well above the TSRC of 2,4-D in rats, and regulatory guidance addressing dose selection for animal bioassays, including the EOGRT, has cautioned that toxicity observed above saturating doses is not relevant for human risk assessment if there is a large disparity between doses reflecting onset of the TSRC compared to real-world human exposures (OECD 2012a, 2012b, 2012c). Second, weakly active non-adverse thyroid effects were observed in the high dose only in pregnant dams in the robust EOGRT study, but importantly, that high dose also was demonstrated to be well above the TSRC in females, and particularly in pregnant females. Third, dog studies of 2,4-D showed no evidence of thyroid toxicity even at lethal doses. In addition, 2,4 D blood concentrations are substantially higher in dogs than rats administered equivalent external doses (van Ravenswaay et al. 2003), primarily because dogs do not clear 2,4-D as efficiently as rats and humans do (Timchalk 2004). Fourth, the recently completed EDSP Tier 1 assays failed to detect any signal of adverse thyroid activity in frogs, in which certain developmental changes are specific for thyroid toxicity. The lack of findings in the dogs and the frog supports the conclusion that the rat is uniquely susceptible to hypothyroidism due to the poor binding of T4 to the carrier proteins in rat blood (Jahnke et al. 2004) making the rat T4 uniquely susceptible to competitive displacement by 2,4-D (van den Berg et al. 1991). Finally, a lack of biological plausibility is further affirmed by the extremely large margin of exposure between biomonitored 2,4-D doses reported for male farm-worker applicators in the Ag Health Study itself and the NOEL dose for thyroid effects in rats reported by Stoker. The Alexander et al. (2007) study of farm families identified a geometric mean exposure dose for male applicators of 2.46 µg/kg/day, which is approximately 10 000X below the NOEL of 30 mg/kg/day (30 000 µg/kg/day) for thyroid effects identified by Stoker and Zorrilla (2010). Importantly, the geometric mean dose for female spouses living in close proximity to active 2,4-D application operations was 0.08 µg/kg/day, and was substantially disparate (~300 000) from the approximately 25 mg/kg/day dietary dose identified as the inflection point for onset of TSRC in female rats (a non-thyroid toxic dose in rats). These large margins of exposures have been confirmed in other high quality biomonitoring studies of farmer-applicators in which a geometric mean dose of 1.6 µg/kg/day was reported (Thomas et al. 2010).

Thus, a WoE evaluation of potential effects of 2,4 D on the HPT axis indicates no concern for a hypothyroid disease or thyroid tumor outcome in humans.

Conclusions

The Tier 1 EDSP studies and the mammalian Tier 2 EDSP equivalent EOGRT dietary toxicity study of 2,4-D are reliable studies and provide a robust basis for assessing interactions of 2,4-D with the estrogen, androgen and steroidogenesis pathways, and the HPT axis. Key conclusions from the WoE evaluation of the EDSP studies and key toxicological studies include:

- 2,4-D clearly does not demonstrate the potential to interact directly with the estrogen pathway in toxicological studies, including an EDSP Tier 2 equivalent mammalian EOGRT dietary toxicity study in which the top dose exceeded the TSRC, a FSTRA tested to the limit concentration, and a quail reproductive toxicity study, or in high quality studies from the published literature. In addition, EDSP Tier 1 *in vitro* assays, high quality published *in vitro* assays, and ToxCast™ *in vitro* screening studies were negative for estrogen pathway interactions.
- 2,4-D does not demonstrate the potential to interact directly with the androgen pathway in toxicological studies, including an EDSP Tier 2 equivalent mammalian EOGRT dietary toxicity study in which the top dose exceeded the

TSRC, a FSTRA tested to the limit concentration and a quail reproductive toxicity study, or in high quality studies from the published literature. In addition, EDSP Tier 1 *in vitro*, high quality published *in vitro* assays and ToxCast™ *in vitro* studies were negative for androgen pathway interactions.

- 2,4-D showed no robust evidence of interaction with the steroidogenesis pathway in an EDSP Tier 2 equivalent mammalian EOGRT study in which the top dose exceeded the TSRC. 2,4-D effects on steroidogenesis parameters in other studies are likely related to high-dose specific systemic toxicity at doses exceeding the TSRC and are not likely to be endocrine mediated.
- 2,4-D showed no adverse interactions with the HPT axis in an EDSP Tier 2 equivalent mammalian EOGRT study in which the top dose exceeded the TSRC. It interacts with the HPT axis in rats (which is clearly a species susceptible to thyroid interactions and not predictive of thyroid effects in other species for compounds acting on the thyroid by the mechanism demonstrated for 2,4-D—displacement of thyroxine from plasma-binding sites) at high doses exceeding the TSRC in mammals and substantially exceeding human systemic doses identified in high quality biomonitoring studies. The thyroid-sensitive AMA tested to the assay limit concentration was negative.
- The EOGRT dietary toxicity study is an acceptable EDSP Tier 2-equivalent mammalian study in which the top dose exceeded the TSRC, and predicts no adverse endocrine-related toxicity to mammals. This study provides a robust basis for concluding that the NOAEL for any endocrine effects is higher than the NOAELs currently used as points of departure for acute, subchronic or chronic human health risk assessment.
- No studies, including high quality studies in the published literature, predict significant endocrine-related toxicity or functional decrements in any species at environmentally relevant concentrations, or, in mammals, at doses below the TSRC that are relevant for human hazard and risk assessment.

Overall, there is no basis for concern regarding a potential for interaction of 2,4-D with endocrine pathways or axes (estrogen, androgen, steroidogenesis, or thyroid), and thus 2,4-D is unlikely to pose a threat from endocrine disruption to wildlife or humans under conditions of real-world exposures. This conclusion is consistent with a similar but less comprehensive WoE review of the 2,4-D endocrine disruption data conducted the US EPA (US EPA 2015), which stated that there was "no convincing evidence of potential interaction [of 2,4-D] with the estrogen, androgen or thyroid pathways." In addition, EPA concluded there was no need for additional EDSP Tier 2 testing given the availability of the EOGRT study that was regarded as equivalent to the EDSP Tier 2 study.

Acknowledgments

The authors thank the four reviewers, selected by the Editor and anonymous to the authors, for their constructive comments that improved its overall quality and clarity of the manuscript.

In memoriam, Barbara Neal, DABT (prepared by James Lamb):

Barbara Neal and I worked together for over 20 years. The response by her colleagues to her passing remind *me* how valued she was in the field of toxicology. Barbara and I evaluated dozens of various issues in reproductive, developmental and endocrine responses. She approached work and life with a sense of humor and passion that will be missed by me and others in toxicology. Very few will ever match her keen observations and care at interpretation.

Barbara would dig into data more deeply than most. She would find interesting and useful information often missed by others. The paper by Barbara in this issue of Critical Reviews in Toxicology is a perfect example of Barbara's passion for paying attention to detail. She has analyzed and understood the vast dataset on 2,4-D in a way that no one else could match.

Barbara was a creative and open-minded toxicologist who always looked for more clear ways to describe data, and stronger methods to test a hypothesis. She had no patience with sloppy or convenient interpretations. She would become annoyed with scientists who worked to prove their own pre-existing views, which she felt had become far too common and too adversarial. Barbara always sought an honest answer to an honest question without malice or some hidden agenda. She had no patience for hiding or overlooking results to prove a point.

Barbara worked with her own unique sense of humor. She enjoyed puns and surprising twists in a story. You could often hear her chuckling in a crowd, often at her own joke. She laughed often, even at herself, but never at the expense of others.

Barbara Neal was a special scientist and person who will be deeply missed by many of us in toxicology, which she called *home*.

Barbara passed away on October 19, 2015, during the final stages of this manuscript preparation. She was a member of the Society of Toxicology and a Diplomate of the American Board of Toxicology, and held the position of Senior Managing Scientist at Exponent, Inc. since 2010. Her career in toxicology extended over 30 years, and included earlier positions at The Weinberg Group, Inc., BBL Sciences, and Battelle Columbus Laboratories.

Declaration of interest

The employment affiliation of the authors is as shown on the cover page. This review was funded by the Industry Task Force II on 2,4-D Research Data [Authors Neal (deceased), Bus, Williams, Staveley and Lamb work for Exponent, which is a consulting company that has performed work for the Industry Task Force II on 2,4-D Research Data, as well as for individual member companies of the Task Force who manufacture 2,4-D. Authors Coady and Marty work for The Dow Chemical Company, which manufactures 2,4-D. On behalf of a previous employer and manufacturer of 2,4-D (The Dow Chemical Company), author Bus has engaged in a single litigation case (defendant deposition).] The review is the exclusive work product of the authors. The professional opinions expressed and the conclusions drawn are those of the authors and not necessarily those of their employers or the sponsors. This review was funded by the Industry Task Force II on 2,4-D Research Data.

Supplemental material

Supplemental material for this article is available online here.

References

Alexander HC, Mayes MA, Gersich FM. 1983. The acute toxicity of (2,4-dichlorophenoxy) acetic acid to representative aquatic organisms. Unpublished study by The Dow Chemical Co., Midland, MI. Lab Report ES-584.

Alexander BH, Mandel JS, Baker BA, Burns CJ, Bartels MJ, Acquavella JF, Gustin C. 2007. Biomonitoring of 2,4 dichlorophenoxyacetic acid exposure and dose in farm families. Environ Health Perspect. 115:370–376.

Aronzon CM, Sandoval MT, Herkovits J, Perez-Coll CS. 2011. Stage-dependent toxicity of 2,4-dichlorophenoxyacetic on the embryonic development of a South American toad, *Rhinella arenarum*. Environ Toxicol. 26:0373–0381

Aylward LL, Hays SM. 2008 Biomonitoring equivalents (BE) dossier for 2,4-dichlorophenoxyacetic acid (2,4-D) (CAS No. 94-75-7). Regul Toxicol Pharmacol. 51(3 Suppl):.

Aylward LL, Morgan MK, Arbuckle TE, Barr DB, Burns CJ, Alexander BH, Hays SM. 2010. Biomonitoring data for 2,4-dichlorophenoxyacetic acid in the United States and Canada: interpretation in a public health risk assessment context using biomonitoring equivalents. Environ Health Perspect. 118:177–181.

Bage G, Cekanova E, Larsson KS. 1973. Teratogenic and embryotoxic of the herbicides di and trichlorophenoxyacetic acids (2,4-D and 2,4,5-T). Acta Pharmacol Toxicol (Copenh). 32:408–416.

Barton HA, Pastoor TP, Baetcke K, Chambers JE, Dilberto J, Doerrer NG, Driver JH, Hastings CE, Iyengar S, Krieger R, et al. 2006. The acquisition and application of absorption, distribution, metabolism, and excretion (ADME) data in agricultural chemical safety assessments. Crit Rev Toxicol. 36:9–35.

Bernal J. 2012. Thyroid hormones in brain development and function. Thyroid Disease Manager. Retrieved from: http://www.thyroidmanager.org/chapter/thyroid hormones-in-brain-development and-function/

Blair RM, Fang H, Branham WS, Haas BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. 2000. The estrogen receptor relative binding affinities of 188 natural and xenochemicals: structural diversity of ligands. Tox Sci. 54:138–153.

Blakley PM, Kim JS, Firneisz GD 1989. Effects of paternal subacute exposure to Tordon 202c on fetal growth and development in CD-1 mice. Teratology. 39:237–241.

Borgert CJ, Stuchal LD, Mihaich EM, Becker RA, Bentley KS, Brausch JM, Coady K, Geter DR, Gordon E, Guiney PD, et al. 2014. Relevance weighting of tier 1 endocrine screening endpoints by rank order. Birth Defects Res B Dev Repro Tox. 101:90–113.

Buist SC, Klaassen CD. 2004. Rat and mouse differences in gender-predominant expression of organic anion transporter (Oat1-3; Slc22a6-8) mRNA levels. Drug Metab Dispos. 32:620–625.

Burns CJ, Swaen GM. 2012. Review of 2,4-dichlorophenoxyacetic acid (2,4-D) biomonitoring and epidemiology. Crit Rev Toxicol 42:768–786.

Bus JS, Hammond LE. 2007. Regulatory progress, toxicology, and public concerns with 2,4-D: where do we stand after two decades?. Crop Protect. 26:266 269.

Carmichael NG, Barton HA, Boobis AR, Cooper RL, Dellarco VL, Doerrer NG, Fenner-Crisp PA, Doe JE, Lamb JC, Pastoor TP. 2006. Agricultural chemical safety assessment: a multi-sector approach to the modernization of human safety requirements. Crit Rev Toxicol. 36:1–7

Carney EW, Zablotny CL, Marty MS, Crissman J, Anderson P, Woolhiser M, Holsapple M. 2004. The effects of feed restriction during *in utero* and postnatal development in rats. Toxicol Sci. 82:237–249.

Cavieres MF, Jaeger J, Potter W. 2002. Developmental toxicity of a commercial herbicide mixture in mice: i. effects on embryo implantation and litter size. Environ Health Perspect. 110:1081–1085.

CDC (Centers for Disease Control and Prevention). 2005. Third national repott on human exposure to environmental chemicals. Atlanta (GA): Centers for Disease Control and Prevention.

Chapin RE, Gulati DK, Barnes LH, Teague JL. 1993. The effects of feed restriction on reproductive function in Sprague-Dawley rats. Fundam Appl Toxicol. 20:23–29.

Charles JM, Cunny HC, Wilson RD, Bus JS. 1996a. Comparative subchronic studies on 2,4-dichlorophenoxyacetic acid, amine and ester in rats. Fundam Appl Toxicol. 33:161–165.

Charles JM, Dalgard DW, Cunny HC, Wilson RD, Bus JS. 1996b. Comparative subchronic and chronic dietary toxicity studies on 2,4-dichlorophenoxyacetic acid, amine, and ester in the dog. Fundam Appl Toxicol. 29:78–85.

Charles JM, Bond DM, Jeffries TK, Yano BL, Stott WT, Johnson KA, Cunny HC, Wilson RD, Bus JS. 1996c. Chronic dietary toxicity/oncogenicity studies on 2,4-dichlorophenoxyacetic acid in rodents. Fundam Appl Toxicol. 33:166–172.

Charles JM, Hanley TR Jr. Wilson RD, van Ravenswaay B, Bus JS. 2001. Developmental toxicity studies in rats and rabbits on 2,4-dichlorophenoxyacetic acid and its forms. Toxicol Sci. 60:121–131.

Clark RL. 1999. Endpoints of reproductive system development. An evaluation and interpretation of reproductive endpoints for human risk assessment. Washington, D.C.: International Life Sciences Institute, Health and Environmental Science Institute; p. 27–62.

Coady KC, Marino TA, Thomas J. 2010. 2,4-dichlorophenoxyacetic acid: the amphibian metamorphosis assay using the South African clawed frog, *Xenopus laevis*. Unpublished study by The Dow Chemical Company, Midland, MI. Study ID: 101025.

Coady KK, Sosinski LK. 2011. 2,4-dichlorophenoxyacetic acid: evaluation of 2,4-dichlorophenoxyacetic acid in the human recombinant aromatase assay. Unpublished study by The Dow Chemical Company, Midland, MI. Study ID 111036.

Coady K, Marino T, Thomas J, Sosinski L, Neal B, Hammond L. 2013. An evaluation of 2,4-dichlorophenoxyacetic acid in the amphibian metamorphosis assay and the fish short-term reproduction assay. Ecotoxicol Environ Saf. 90:143–150.

Coady KC, Kan LL, Schisler MR, Gollapudi BB, Neal B, Williams A, LeBaron M. 2014. Evaluation of potential endocrine activity of 2,4 dichlorophe noxyacetic acid using *in vitro* assays. Toxicol In Vitro. 28:1018–1025.

Collins TF, Williams CH. 1971. Teratogenic studies with 2,4,5-T and 2,4-D in the hamster. Bull Environ Contam Toxicol. 6:559–567.

Cooper RL, Lamb JC, Barlow SM, Bentley K, Brady AM, Doerrer NG, Eisenbrandt DL, Fenner-Crisp PA, Hines RN, Irvine L, et al. 2006. A tiered approach to life stages testing for agricultural chemical safety assessment. Crit Rev Toxicol. 36:69–98.

Cox LA, Popken D, Marty MS, Rowlands JC, Patlewicz G, Goyak KO, Becker RA. 2014. Developing scientific confidence in HTS-derived prediction models: lessons learned from an endocrine case study. Regul Toxicol Pharmacol. 69:443–450.

Crain DA, Guillette LJ Jr. Rooney AA, Pickford DB. 1997. Alterations in steroidogenesis in alligators (*Alligator mississippiensis*) exposed naturally and experimentally to environmental contaminants. Environ. Health Perspect. 105:528–533.

Crain DA, Spiteri ID, Guillette LJ Jr. 1999. The functional and structural observations of the neonatal reproductive system of alligators exposed *in ovo* to atrazine, 2,4-D, or estradiol Toxicol Ind Health. 15:180–185.

Dalgard DW. 1993a. 13-Week dietary study with 2,4-D in dogs. Unpublished study by Hazleton Washington Inc. Project No. HWA 2184-125.

Dalgard DW. 1993b. 52-Week dietary study with 2,4-D in dogs. Unpublished study by Hazleton Washington Inc. Project No. HWA 2184-124.

De Peyster A, Mihaich E. 2014. Hypothesis-driven weight of evidence analysis to determine potential endocrine activity of MT8E. Regul Toxicol Pharmacol. 69:348–370.

Dinamarca VM, Hidalgo ME, Cavieres MF. 2007. Lack of effects of 2,4-dichlorophenoxyacetic acid administration on markers of oxidative stress during early pregnancy in mice. Toxicology. 237:104–110.

Dinse GE, Peddada SD, Harris SF, Elmore SA. 2010. Comparison of historical control tumor incidence rats in female Harlan Spraque-Dawley and Fischer 344/N rats from two-year bioassays performed by the National Toxicology Program. Toxicol Path. 38:765–775.

Dohler KD, Wong CC, von zur Muhlen A. 1979. The rat as model for the study of drug effects on thyroid function: consideration of methodological problems. Pharmacol Ther B. 5:305–318.

Duffard R, Garcia G, Rosso S, Bortolozzi A, Madariaga M, di Paolo O, Evangelista de Duffard AM. 1996. Central nervous system myelin deficit in rats exposed to 2,4-dichlorophenoxyacetic acid throughout lactation. Neurotoxicol Teratol. 18:691–696.

Edwards DF, Kay R. 1985. Weight and vaginal opening in the albino rat. J Reprod Fertil. 73:1–8.

EFSA (European Food Safety Authority). 2014. Conclusion on pesticide peer review. Conclusion on the peer review of the pesticide risk assessment of the active substance 2,4-D. EFSA J. 12:3812.

Elo HA, MacDonald E. 1989. Effects of 2,4-dichlorophenoxyacetic acid (2,4-D) on biogenic amines and their acidic metabolites in brain and cerebrospinal fluid of rats. Arch Toxicol. 63:127–130.

Fang H, Tong W, Branham WS, Moland CL, Dial SL, Hong H, Xie Q, Perkins R, Owens W, Sheehan DM. 2003. Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. Chem Res Toxicol. 16:1338–1358.

Fliers E, Alkemade A, Wiersinga WM, Swaab DF. 2006. Hypothalamic thyroid hormone feedback in health and disease. Prog Brain Res. 153:189–207.

Florsheim WH, Velcoff SM. 1962. Some effects of 2,4-dichlorophenoxyacetic acid on thyroid function in the rat: effects on iodine accumulation. Endocrinology. 71:1–6.

Florsheim WH, Velcoff SM, Williams AD. 1963. Some effects of 2,4-dichlorophenoxyacetic acid on thyroid function in the rat: effect on peripheral thyroxine . Endocrinology. 72:327 333.

Garabrant DH, Philbert MA. 2002. Review of 2,4 dichlorophenoxyacetic acid (2,4-D) epidemiology and toxicology. Crit Rev Toxicol. 32:233 257

Garry VF, Tarone IR, Kirsch JM, Abdallah DP, Lombardi LK, Long BL, Burroughs DB, Kesner JS. 2001. Biomarker correlations of urinary 2,4-D levels in foresters: genomic instability and endocrine disruption. Environ Health Perspect. 109:495–500.

Goedken MJ, Kerlin RL, Morton D. 2008. Spontaneous and age related testicular findings in beagle dogs. Toxicol Pathol. 36:465–471.

Goldey ES, Kehn LS, Rehnberg GL, Crofton KM. 1995. Effects of develop mental hypothyroidism on auditory and motor function in the rat. Toxicol Appl Pharmacol. 135:67–76.

Goldner WS, Sandler DP, Yu F, Shostrom V, Hoppin JA, Kamel F, Levan TD. 2013. Hypothyroidism and pesticide use among male private pesticide applicators in the Agricultural Health Study. J Occup Environ Med. 55:1171–1178.

Goldner WS, Sandler DP, Yu F, Hoppin JA, Kamel F, Levan TD. 2010. Pesticide use and thyroid disease among women in the Agricultural Health Study. Am J Epidemiol. 171:455–464.

Gorzinski SJ, Kociba RJ, Campbell RA, Smith FA, Nolan RJ, Eisenbrandt DL. 1987. Acute, pharmacokinetic, and subchronic toxicological studies of 2,4-dichlorophenoxyacetic acid. Fundam Appl Toxicol. 9:423–435.

Gorzinski SJ, Wade CE, Morden DC, Keyes DG, Wolfe EL, Dittenber DA, Kalnins RV, Schuetz DJ, Kociba RJ. 1981a. Technical grade 2,4-dichlorophenoxyacetic acid (2,4-D): results of a 13-week subchronic dietary toxicity study in the CDF Fischer 344 rat. Unpublished study, The Dow Chemical Company, Midland, MI.

Gorzinski SJ, Wade CE, Morden DC, Keyes DG, Dittenber DA, Kalnins RV, Schuetz DJ, Kociba RJ. 1981b. Purified 2,4-dichlorophenoxyacetic acid (2,4-D): results of a 13-week subchronic dietary toxicity study in the CDF Fischer 344 Rat. Unpublished study, The Dow Chemical Company, Midland, MI.

Hasegawa M, Kusuhara H, Endou H, Sugiyama Y. 2003. Contribution of organic anion transporters to the renal uptake of anionic compounds and nucleoside derivatives in rat. J Pharmacol Exp Ther. 305:1087–1097.

Hassanein K. 2012. Histopathological effects of 2,4-dichlorophenoxyacetic acid on Sprague Dawley rats with special reference to its possible carcinogenicity. Vet World. 5:24–30.

Hays SM, Aylward LL, Driver J, Ross J, Kirman C. 2012. 2,4-D exposure and risk assessment: comparison of external dose and biomonitoring based approaches. Regul Toxicol Pharmacol. 64:481–489.

Health Canada. 2010. Report on human biomonitoring of environmental chemicals in Canada. http://www.hc-sc.gc.ca/ewh-semt/pubs/contaminants/chms-ecms/index-eng.php

Hecker M, Giesy JP, Timm G. 2008. Multi-laboratory validation of the H295R steroidogenesis assay to identify modulators of testosterone and estradiol production. Report by ENTRIX, Inc., submitted to US-EPA, Contract Code # GS-10F-0041L.

Heggstrom MJ. 2009. The sublethal effects of 2,4-D dimethylamine on wood frog tadpoles in Saskatchewan. A thesis submitted to the College of Graduate Studies and Research for the Degree of Master of Science in the Toxicology Graduate Program, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.

Helmreich DL, Tylee D. 2011. Thyroid hormone regulation by stress and behavioral differences in adult male rats. Horm Behav. 60:284–291.

Hill RN, Erdreich LS, Paynter OE, Roberts PA, Rosenthal SL, Wilkinson CF. 1989. Thyroid follicular cell carcinogenesis. Fundam Appl Toxicol. 12:629–697.

Hoberman AM. 1990. Developmental toxicity (embryo-fetal toxicity and teratogenic potential) study of 2,4-dichlorophenoxyacetic acid (2,4-D) administered orally via stomach tube to New Zealand white rabbits. Unpublished study, Argus Research Laboratories, Inc., Horsham, PA. Project No. 320-003.

Holcombe GW, Benoit DA, Hammermeister DE, Leonard EN, Johnson RD. 1995. Acute and long-term effects of nine chemicals on the Japanese medaka (*Oryzias latipes*). Arch Environ Contam Toxicol. 28:287–297.

Hotchkiss AK, Parks-Saldutti LG, Ostby JS, Lambright C, Furr J, Vandenbergh JG, Gray LE Jr. 2004. A mixture of the "antiandrogens" linuron and butyl benzyl phthalate alters sexual differentiation of the male rat in a cumulative fashion. Biol Reprod. 71:1852–1861.

Jahnke GD, Choksi NY, Moore JA, Shelby MD. 2004. Thyroid toxicants: assessing reproductive health effects. Environ Health Perspect. 112:363–368.

Jeffries TK, Yano BL, Ormand JR, Battjes JE. 1995. 2,4-dichlorophenoxyacetic acid: chronic toxicity/oncogenicity study in Fischer 344 rats. Unpublished study, The Dow Chemical Company, Midland, MI. Project No. K 002372-064F.

Judson RS, Houck KA, Kavlock RJ, Knudsen TB, Martin MT, Mortensen HM, Reif DM, Rotroff DM, Shah I, Richard AM, Dix DJ. 2010. *In vitro* screening of environmental chemicals for targeted testing prioritization – the ToxCast™ project. Environ Health Persp. 118:485–492.

Jung J, Ishida K, Nishihara T. 2004. Anti-estrogenic activity of fifty chemicals evaluated by *in vitro* assays. Life Sci. 74:3065–3074.

Jungbauer A, Beck V. 2002. Yeast reporter system for rapid determination of estrogenic activity. J Chromatogr B Analyt Technol Biomed Life Sci. 777:167–178.

Kim CS, Keizer RF, Pritchard JB. 1988. 2,4-Dichlorophenoxyacetic acid intoxication increases its accumulation within the brain. Brain Res. 440:216–226.

Kim HC, Kim WD, Kwon TH, Kim DH, Park YI, Dong MS. 2002. Mechanism of phenoxy compounds as an endocrine disruptor. J Toxicol Pub Health. 18:331–339.

Kim HJ, Park YI, Dong MS. 2005. Effects of 2,4-D and DCP on the DHT-induced androgenic action in human prostate cancer cells. Toxicol Sci. 88:52–59.

Kimmel CA, Garry MR, DeSesso JM. 2014. Relationship between bent long bones, bent scapulae, and wavy ribs: malformations or variations?. Birth Defects Res B Dev Reprod Toxicol. 101:379–392.

Klimisch HJ, Andreae M, Tillmann U. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. Regul Toxicol Pharmacol. 25:1–5.

Kobal S, Cebulj-Kadunc N, Cestnik V. 2000. Serum T3 and T4 concentrations in the adult rats treated with herbicide 2,4-dichlorophenoxyacetic acid. Pflugers Arch. 440:R171–R172.

Koç ND, Akbulut C. 2012. Histological analysis of acute toxicity of 2,4-diclorophenoxyacetic acid in ovary of zebrafish. Anim Cells Syst. 16:400–407.

Kojima H, Katsura E, Takeuchi S, Niiyama K, Kobayashi K. 2004. Screening for estrogen and androgen receptor activities in 200 pesticides by in vitro reporter gene assays using Chinese hamster ovary cells. Environ Health Perspect. 112:524–531.

Knopp D. 1994. Assessment of exposure to 2,4-dichlorophenoxyacetic acid in the chemical industry: results of a five year biological monitoring study. Occup Environ Med. 51:152–159.

LaChapelle AM, Ruygrok ML, Toomer M, Oost JJ, Monnie ML, Swenson JA, Compton AA, Stebbins-Boaz B. 2007. The hormonal herbicide, 2,4-dichlorophenoxyacetic acid, inhibits *Xenopus* oocyte maturation by targeting translational and post-translational mechanisms. Reprod Toxicol. 23:20–31.

Lamb JC, Marks TA, Gladen BC, Allen JW, Moore JA. 1981a. Male fertility, sister chromatid exchange, and germ cell toxicity following exposure to mixtures of chlorinated phenoxy acids containing 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Toxicol Environ Health. 8:825–834

Lamb JC, Moore JA, Marks TA, Haseman JK. 1981b. Development and viability of offspring of male mice treated with chlorinated phenoxy acids and 2,3,7,8-tetrachlorodibenzo-p-dioxin. J Toxicol Environ Health. 8:835–844.

Lamb JC, Marks TA, McConnell EE, Abeywickrama K, Moore JA. 1981c. Toxicity of chlorinated phenoxy acids in combination with 2,3,7,8 tetrachlorodibenzo-p-dioxin in C57BL/6 male mice. J Toxicol Environ Health. 8:815–824.

LeBaron MJ, Kan HL. 2011. Evaluation of 2,4-dichlorophenoxyacetic acid (2,4-D) in an *in vitro* estrogen receptor transcriptional activation assay in human cell line HeLa 9903. Unpublished study report of The Dow Chemical Company, Midland, Michigan 48674. Study ID 111043.

LeBaron MJ, Schisler MR, Visconti NR. 2011a. Evaluation of 2,4-dichloro phenoxy acetic acid (2,4-D) in an *in vitro* estrogen receptor binding assay Unpublished study report of The Dow Chemical Company Midland, Michigan 48674. Study ID: 111121.

LeBaron MJ, Schisler MR, Visconti NR. 2011b. Evaluation of 2,4-dichlorophenoxyacetic acid (2,4-D) in an *in vitro* androgen receptor binding assay. Unpublished study report of The Dow Chemical Company, Midland, Michigan, 48674. Study ID 111111.

LeBaron MJ, Kan HL, Perala AW. 2011c. Evaluation of 2,4-dichlorophenoxyacetic acid (2,4-D) in the *in vitro* steroidogenesis assay. Unpublished study report of The Dow Chemical Company, Midland, Michigan 48674. Study ID 111038.

Lee HS, Sasagawa SI, Kato S, Fukuda R, Horiuchi H, Ohta A. 2006. Yeast two-hybrid detection systems that are highly sensitive to a certain kind of endocrine disruptors. Biosci Biotechnol Biochem. 70:521–524.

Lemaire G, Mnif W, Mauvais P, Balaguer P, Rahmani R. 2006. Activation of alpha- and beta-estrogen receptors by persistent pesticides in reporter cell lines . Life Sci. 79:1160–1169.

Lerda D, Rizzi R. 1991. Study of reproductive function in persons occupationally exposed to 2,4-dichlorophenoxyacetic acid (2,4-D). Mutat Res. 262:47–50.

Lin N, Garry VF. 2000. *In vitro* studies of cellular and molecular developmental toxicity of adjuvants, herbicides, and fungicides commonly used in Red River Valley, Minnesota. J Toxicol Environ Health Part A. 60:423–439.

Ljubojevic M, Herak-Kramberger CM, Hagos Y, Bahn A, Endou H, Burckhardt G, Sabolic I. 2004 Rat renal cortical OAT1 and OAT3 exhibit gender differences determined by both androgen stimulation and estrogen inhibition. Am J Physiol Renal Physiol. 287: F124–F138.

Marino TA, Coady KK, Sosinski LK, Thomas J. 2010. 2,4-dichlorophenoxyacetic acid: a fish short term reproduction assay using the Fathead Minnow, *Pimephales promelas*. Unpublished study report of The Dow Chemical Company Study ID: 101026.

Marty MS, Crissman JW, Carney EW. 2001. Evaluation of the male pubertal assay's ability to detect thyroid inhibitors and dopaminergic agents. Toxicol Sci. 60:63–76.

Marty MS, Johnson KA, Carney EW. 2003. Effect of feed restriction on Hershberger and pubertal male assay endpoints. Birth Defects Res B Dev Reprod Toxicol. (4), 363–374.

Marty MS, Zablotny CL, Andrus AK, Boverhof DR, Bus JS, Perala AW, Saghir S, Yano BL. 2010. 2,4-D: an extended one-generation dietary toxicity study in Crl:CD(SD) rats. Unpublished study report of The Dow Chemical Company, Midland, Michigan.

Marty MS, Neal BH, Zablotny CL, Yano BL, Andrus AK, Woolhiser MR, Boverhof DR, Saghir SA, Perala AW, Passage JW, et al. 2013. An F1-extended one-generation reproductive toxicity study in Crl:CD(SD) rats with 2,4-dichlorophenoxyacetic acid Toxicol Sci. 136:527–547.

Mayes MA, Gorzinski SJ, Potter RB, Richardson CH. 1990. 2,4-dichlorophenoxyacetic acid: evaluation of the toxicity to early life stages of the fathead minnow, *Pimephales promelas* rafinesque. Unpublished study by The Dow Chemical Company. Lab Report ES-2157.

McIntyre BS, Barlow NJ, Foster PMD. 2001. Androgen-mediated development in male rat offspring exposed to flutamide in utero: permanence and correlation of early postnatal changes in anogenital distance and nipple retention with malformations in androgen-dependent tissues. Toxicol Sci. 62:236–249.

Mitchell LR, Beavers JB, Martin KH, Jaber M. 2000. 2,4-D acid: a reproduction study with the northern bobwhite. Unpublished study by Wildlife International, Ltd, Lab Project Number: 467-106.

Morgan MK, Scheuerman PR, Bishop CS, Pyles RA. 1996. Teratogenic potential of atrazine and 2,4-D using FETAX. J Toxicol Environ Health. 48:151–168.

Munro IC, Carlo GL, Orr JC, Sund KG, Wilson RM, Kennepohl E, Lynch BS, Jablinski M, Lee NL. 1992. A comprehensive integrated review of the scientific evidence relating to the safety of the herbicide 2,4-D. J Am Coll Toxicol. 11:559–664.

Nieuwkoop PD, Faber J. 1994. Normal table of xenopus laevis (daudin): a systematical and chronological survey of the development from the fertilized egg till the end of metamorphosis. New York (NY): Garland Publishing, Inc.

Nishihara T, Nishikawa JI, Kanayama T, Dakeyama F, Saito K, Imagawa M, Takatori S, Kitagawa Y, Hori S, Utsumi H. 2000. Estrogenic activities of 517 chemicals by yeast two-hybrid assay. J Health Sci. 46:282–298.

Oakes DJ, Webster WS, Brown-Woodman PD, Ritchie HE. 2002a. A study of the potential for a herbicide formulation containing 2,4-D and picloram to cause male-mediated developmental toxicity in rats. Toxicol Sci. 68:200–206.

Oakes DJ, Webster WS, Brown-Woodman PD, Ritchie HE. 2002b. Testicular changes induced by chronic exposure to the herbicide formulation, Tordon 75D (2,4-dichlorophenoxyacetic acid and picloram) in rats. Reprod Toxicol. 16:281–289.

Obidike I, Aka L, Omoja V, Shoyinka S, Kamalu T, Anika S. 2012. Testicular morphology and antispermatogenic effects of 2,4-dichlorophenoxyacetic acid (2,4-D) in male West African Dwarf (WAD) goats. Comp Clin Path. 21:1457–1462.

O'Connor CM, Gilmour KM, Van Der Kraak G, Cooke SJ. 2011. Circulating androgens are influenced by parental nest defense in a wild teleost fish. J Comp Physiol a Neuroethol Sens Neural Behav Physiol. 197:711–715.

Odum J, Tittensor S, Ashby J. 1998. Limitations of the MCF-7 cell proliferation assay for detecting xenobiotic oestrogens. Toxicol In Vitro. 12:273–278.

OECD. 2012a. Test No. 443: extended one-generation reproductive toxicity study, OECD guidelines for the testing of chemicals, Section 4. Paris: OECD Publishing.

OECD 2012b. Guidance document on standardized test guidelines for evaluating chemicals for endocrine disruption. Series on testing and assessment, No. 150, ENV/JM/MONO(2012)22, 24-Aug-2012.

OECD. 2012c. Guidance document 116 on the conduct and design of chronic toxicity and carcinogenicity studies, supporting test guidelines 451, 452, and 453. 2nd Ed., 13-Apr-2012.

Orr TE, Mann DR. 1990. Effects of restraint stress on plasma LH and testosterone concentrations, Leydig cell LH/hCG receptors, and in vitro testicular steroidogenesis in adult rats. Horm Behav. 24:324–341.

Orr TE, Taylor MF, Bhattacharyya AK, Collins DC, Mann DR. 1994. Acute immobilization stress disrupts testicular steroidogenesis in adult male rats by inhibiting the activities of 17 alpha-hydroxylase and 17,20-lyase without affecting the binding of LH/hCG receptors. J Androl. 15:302–308.

Orton F, Lutz I, Kloas W, Routledge EJ. 2009. Endocrine disrupting effects of herbicides and pentachlorophenol: in vitro and in vivo evidence. Environ Sci Technol. 43:2144–2150.

Ottinger MA, Abdelnabi M, Quinn M, Golden N, Wu J, Thompson N. 2002. Reproductive consequences of EDCS in birds. What do laboratory effects mean in field species? Neurotox Teratol. 24:17–28.

Padilla S, Corum D, Padnos B, Hunter DL, Beam A, Houck KA, Sipes N, Kleinstreuer N, Knudsen T, Dix DJ, Reif DM. 2012. Zebrafish developmental screening of the ToxCast™ Phase I chemical library. Reprod Toxicol. 33:174–187.

Partsch CJ, Abrahams S, Herholz N, Peter M, Veldhuis JD, Sippell WG. 1994. Variability of pulsatile luteinizing hormone secretion in young male volunteers. Eur J Endocrinol. 131:263–272.

Payne J, Jones C, Lakhani S, Kortenkamp A. 2000. Improving the reproducibility of the MCF-7 cell proliferation assay for the detection of xenoestrogens. Sci Total Environ. 248:51–62.

Petit F, Le Goff P, Cravedi JP, Valotaire Y, Pakdel F. 1997. Two complementary bioassays for screening the estrogenic potency of xenobiotics; recombinant yeast for trout estrogen receptor and trout hepatocyte cultures. J Mol Endocrinol. 19:321–335.

PMRA (Pesticide Management Regulatory Authority). 2007 Proposed acceptability for continuing registration. Re-evaluation of the agricultural, forestry, aquatic and industrial site uses of (2,4-dichlorophenoxy)acetic acid [2,4-D]. PACR2007-06, 19 June.

Pochettino AA, Bongiovanni B, Duffard RO, Evangelista de Duffard AM. 2013. Oxidative stress in ventral prostate, ovary, and breast by 2,4-dichlorophenoxyacetic acid in pre- and postnatal exposed rats. Environ Toxicol. 28:1–10.

Rawlings NC, Cook SJ, Waldbillig D. 1998. Effects of the pesticides carbofuran, chlorpyrifos, dimethoate, lindane, triallate, trifluralin, 2,4-D, and pentachlorophenol on the metabolic endocrine and reproductive endocrine system in ewes. J Toxicol Environ Health Part A. 54:21–36.

Rehwold RE, Kelley E, Mahoney M. 1977 Investigations into the acute toxicity and some chronic effects of selected herbicides and pesticides on several fresh water fish species. B Environ Contam Toxicol. 18:361–365.

Reif DM, Martin MT, Tan SW, Houck KA, Judson RS, Richard AM, KnudsenTB, Dix DJ, Kavlock RJ. 2010. Endocrine profiling and prioritization of environmental chemicals using ToxCast™ data. Environ Health Perspect. 118:1714–1720.

Relyea RA. 2005. The impact of insecticides and herbicides on the biodiversity and productivity of aquatic communities. Ecol Appl. 15:618–627

Rodwell DE. 1983. A teratology study in Fischer 344 rats with 2,4 dichlorophenoxyacetic acid. Unpublished study by WIL Research Laboratories, Inc., Ashland, OH. Project No. WIL 81135.

Rodwell DE, Brown WR. 1985. A dietary two-generation reproduction study in Fischer 344 rats with 2,4-dichlorophenoxyacetic acid Unpublished study, WIL Laboratories, Ashland OH. Project No. WIL-81137.

Rotroff DM, Dix DJ, Houck KA, Knudsen TB, Martin MT, McLaurin KW, Reif DM, Crofton KM, Singh AV, Xia M, et al. 2013. Using *in vitro* high throughput screening assays to identify potential endocrine-disrupting chemicals. Environ Health Perspect. 121:7–14.

Saghir SA, Mendrala AL, Bartels MJ, Day SJ, Hansen SC, Sushynski JM, Bus JS. 2006. Strategies to assess systemic exposure of chemicals in subchronic/chronic diet and drinking water studies. Toxicol Appl Pharm. 211:245–260.

Saghir SA, Zablotny CL, Bus JS, Marty MS, Perala AW, Yano BL. 2008a. A dietary dose range-finding and pharmacokinetic study of 2,4-dichlorophenoxyacetic acid (2,4-D) in the pregnant Crl:CD(SD) rat and its offspring in preparation for a subsequent F1-extended one-generation toxicity study in rats unpublished report of The Dow Chemical Company, Midland, Michigan. Report Number HET K-002372-130.

Saghir SA, Perala AW, Clark AJ. 2008b. A dietary titration study of 2,4-dichlorophenoxyacetic acid (2,4-D) pharmacokinetics in female CRL:CD(SD) rats. Laboratory Project Study ID 071210. Unpublished report of Toxicology & Environmental Research and Consulting, The Dow Chemical Company, Midland, Michigan.

Saghir SA, Bartels MJ, Rick DL, McCoy AT, Rasoulpoura RJ, Ellis-Hutchings RG, Marty MS. 2012. Assessment of diurnal systemic dose of agrochemicals in regulatory toxicity testing - an integrated approach without additional animal use Regul Toxicol Pharmacol. 63:321–332.

Saghir SA, Marty MS, Zablotny C, Passage J, Perala A, Neal BH, Hammond L, Bus JS. 2013. Life stage-, sex-, and dose-dependent dietary toxicokinetics and relationship to toxicity of 2,4-dichlorophenoxyacetic acid (2,4-D) in rats: implications for toxicity test dose selection, design, and interpretation. Toxicol Sci. 136:294–307.

Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A, Hartung T, Hoffmann S. 2009. "ToxRTool", a new tool to assess the reliability of toxicological. Toxicol Lett 189:138–144.

Schulze GE. 1990. Subchronic toxicity study in dogs with 2,4 dichlorophenoxyacetic acid. Unpublished study by Hazelton Laboratories America, Inc. Project Number HWA 2184-115.

Schulze GE. 1991a. Subchronic toxicity study in rats with 2,4-dichlorophenoxyacetic acid. Unpublished study by Hazelton Laboratories America, Inc. Project No. 2184-116.

Schulze GE. 1991b. Subchronic toxicity study in mice with 2,4-dichlorophenoxyacetic acid. Unpublished study by Hazelton Laboratories America, Inc. Project No. 2184-117.

Shibutani M, Woo GH, Fujimoto H, Saegusa Y, Takahashi M, Inoue K, Hirose M, Nishikawa A. 2009. Assessment of developmental effects of hypothyroidism in rats from in utero and lactation exposure to antithyroid agents. Reprod Toxicol. 28:297–307.

Somers JD, Moran ET, Reinhart BS. 1974. Effect of external application of pesticides to the fertile egg on hatching success and early chick performance 3. Consequences of combining 2,4-D with picloram and extremes in contamination. B Environ Contam Tox. 11:511–516.

Somers JD, Moran ET, Reinhard BS. 1978a. Reproductive success of hens and cockerels originating from eggs sprayed with 2,4-D, 2,4,5-T and picloram followed by early performance of their progeny after a comparable *in ovo* exposure. B Environ Contam Tox. 20:111–119.

Somers JD, Moran ET, Reinhart BS. 1978b. Hatching success and early performance of chicks from eggs sprayed with 2,4-D, 2,4,5,-T and picloram at various stages of embryonic development. B Environ Contam Toxicol. 20:289–293.

Soto AM, Sonnenschein C, Chung KL, Fernandez MF, Olea N, Serrano FO. 1995. The E-SCREEN assay as a tool to identify estrogens: an update on estrogenic environmental pollutants. Environ Health Persp. 103 (Suppl 7):113–122.

Spiteri ID, Guillette LJ Jr, Crain DA. 1999. The functional and structural observations of the neonatal reproductive system of alligators exposed *in ovo* to atrazine, 2,4-D, or estradiol. Toxicol Ind Health. 15:181–186.

Stebbins Boaz B, Fortner K, Frazier J, Piluso S, Pullen S, Rasar M, Reid W, Sinclair K, Winger E. 2004. Oocyte maturation in *Xenopus laevis* is blocked by the hormonal herbicide, 2,4-dichlorophenoxy acetic acid. Mol Reprod Dev. 67:233–242.

Stoker T, Kaydos E, Jeffrey S, Cooper R. 2007. Effect of 2,4-D exposure on pubertal development and thyroid function in the male Wistar rat. Biol Reprod. 77:75–75.

Stoker TE, Zorrilla LM. 2010. The effects of endocrine disrupting chemicals on pubertal development in the rat: use of the EDSP pubertal assays as a screen, in endocrine toxicology. Eldridge, JC, Stevens, JT, eds., New York, Informa Healthcare, pp. 27–81.

Stott W. 1995a. 2,4 dichlorophenoxyacetic acid: dietary oncogenicity study in B6C3F1 mice two-year final report (female). Unpublished study by the Dow Chemical Company, Midland, MI. Lab Project Number K-002372-063F.

Stott W. 1995b. 2,4-dichlorophenoxyacetic acid: dietary oncogenicity study in male B6C3F1 mice two-year final report. Unpublished study by the Dow Chemical Company, Midland, MI. Lab Project Number K-002372-063MF.

Stürtz N, Evangelista de Duffard AM, Duffard R. 2000. Detection of 2,4-dichlorophenoxyacetic acid (2,4-D) residues in neonates breast-fed by 2,4-D exposed dams. Neurotoxicology. 21:147–154.

Stürtz N, Bongiovanni B, Rassetto M, Ferri A, Evangelista de Duffard AM, Duffard R. 2006. Detection of 2,4-dichlorophenoxyacetic acid in rat milk of dams exposed during lactation and milk analysis of their major components. Food Chem Toxicol. 44:8–16.

Sturtz N, Deis RP, Jahn GA, Duffard R, Evangelista de Duffard AM. 2008. Effect of 2,4-dichlorophenoxyacetic acid on rat maternal behavior. Toxicology. 247:73–79.

Stürtz N, Jahn GA, Deis RP, Rettori V, Duffard RO, Evangelsita de Duffard AM. 2010. Effect of 2,4-dichlorophenoxyacetic acid on milk transfer to the litter and prolactin release in lactating rats. Toxicology. 271:13–20.

Sun H, Chaozong S, Bian Q, Chen X, Chen L, Wange X. 2012. Developing *in vitro* reporter gene assays to assess the hormone receptor activities of chemicals frequently detected in drinking water. J Appl Toxicol. 32:635–641

Swan SH, Kruse RL, Liu F, Barr DB, Drobnis EZ, Redmon JB, Wang C, Brazil C, Overstreet JW. 2003. Study for Future Families Research Group: semen quality in relation to biomarkers of pesticide exposure. Environ Health Perspect. 111:1478–1484.

Thomas KW, Dosemeci M, Coble JB, Hoppin JA, Sheldon LS, Chapa G, Croghan CW, Jones PA, Knott CE, Lynch CF, et al. 2010. Assessment of a pesticide exposure intensity algorithm in the agricultural health study. J Expo Sci Environ Epidemiol. 20:559–569.

Timchalk C. 2004. Comparative inter-species pharmacokinetics of phenoxy acid herbicides and related organic acids: evidence that the dog is not a relevant species for evaluation human health risk. Crit Rev Toxicol. 200:1–19.

US Department of Agriculture Forest Service (USDA). 1998. 2,4-Dichlorophenoxyacetic acid formulations – Human health and ecological risk assessment-final report. Contract #53-3187-5-12, submitted by Syracuse Environmental Research Associates.

US Environmental Protection Agency (EPA). 2005. Reregistration eligibility decision for 2,4-D. EPA 738-R-05-02 (RED).

US Environmental Protection Agency (EPA). 2009a. Endocrine disruptor screening program test guidelines. OPPTS 890.1250: Estrogen receptor binding assay using rat uterine cytosol (ER RUC). EPA 740-C-09-005.

US Environmental Protection Agency (EPA). 2009b. Endocrine disruptor screening program test guidelines. OPPTS 890.1300: Estrogen receptor transcriptional activation [human cell line (HeLa-9903)]. EPA 740-C-09-006.

US Environmental Protection Agency (EPA). 2009c. Endocrine disruptor screening program test guidelines. OPPTS 890.1150: androgen receptor binding (rat prostate cytosol). EPA 640-C-09-003.

US Environmental Protection Agency (EPA). 2009d. Endocrine disruptor screening program test guidelines. OPPTS 890.1200: aromatase (human recombinant). EPA 740-C-09-004.

US Environmental Protection Agency (EPA). 2009e. Endocrine disruptor screening program test guidelines. OPPTS 890.1550: steroidogenesis (human cell line-H294R). EPA 640-C-09-003.

US Environmental Protection Agency (EPA). 2009f. Endocrine disruptor screening program test guidelines. OPPTS 890.1100: amphibian metamorphosis (frog). EPA 740-C-09-002.

US Environmental Protection Agency (EPA). 2009g. Endocrine disruptor screening program test guidelines. OPPTS 890.1350: fish short-term reproduction assay. EPA 740-C09-007.

US Environmental Protection Agency (EPA). 2009h. Endocrine disruptor screening program test guidelines. OPPTS 890.1500: Pubertal developmental and thyroid function intact juvenile/peripubertal male rats. EPA 740-C-09-012.

US Environmental Protection Agency (EPA). 2010. ToxCast™ program summary and primary data. Retrieved at: http://www.epa.gov/ncct/toxcast/index.html.

US Environmental Protection Agency (EPA). 2011. Endocrine disruptor screening program (EDSP), weight of evidence: evaluating results of EDSP tier 1 screening to identify the need for tier 2 testing. EPA-HQ-OPPT-2010-0877. Retrieved at: http://www.regulations.gov/#!docketDetail;D=EPA-HQ-OPPT-2010-0877.

US Environmental Protection Agency (EPA). 2013. Memorandum: transmittal of meeting minutes of the FIFRA Scientific Advisory Panel meeting held July 30–August 1, 2013 on Weight-of-evidence: evaluating results of EDSP Tier 1 screening. October 30, 2013.

US Environmental Protection Agency (EPA). 2015. Memorandum. EDSP weight of evidence conclusions on the tier 1 screening assays for the List 1 chemicals. EDSP: weight of evidence analysis of potential interaction with estrogen, androgen or thyroid pathways, Chemical: 2,4-dichlorophenoxy acetic acid (2,4-D). June 29, 2015. Last accessed at: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.epa.gov_sites_production_files_2015-2D06 documents 24-2Dd-2D030001-5F2015-2D06-2D29-5Ftxr0057151.pdf&d=CwIFaQ&c=t0wRGL5ICVzH157W8C8Wew&r=PM43qah _quXCXHJ2v8MievBc17ibr-tvvVsnfoH2TU&m oYJsD_Z4cwpF0N09iL8QLW14EOCrJLD7RkR-IohLzgk&s -TCdmI2pJWbO0MoRsG1tTjnas2Y814eH2ZGecGxhIbA&e

US Environmental Protection Agency (EPA). 2016. Use of high throughput assays and computational tools in the endocrine disruptor screening program. Last accessed at: https://urldefense.proofpoint.com/v2/url?u=https-3A__www.epa.gov_endocrine-2Ddisruption_use-2Dhigh-2Dthroughput-2Dassays-2Dand-2Dcomputational-2Dtools-2Dendocrine-2Ddisruptor&d CwIFaQ&c=t0wRGL5ICVzH157W8C8Wew&r=PM43qah-_quXCXHJ2v8MievBc17ibr-tvvVsnfoH2TU&m oYJsD Z4cwpF0N09iL8QLW14EOCrJLD7RkR-IohLzgk&s=_2ZqJ5xHdbFJ0vwDn9eSFhj_GtgVvnpKJBGqyEryCTI&e

Van den Berg KJ, van Raaij JA, Bragt PC, Notten WR. 1991. Interactions of halogenated industrial chemicals with transthyretin and effects on thyroid hormone levels *in vivo*. Arch Toxicol. 65:15–19.

Van Ravenswaay B, Hardwick TD, Needham D, Pethen S, Lappin GJ. 2003. Comparative metabolism of 2,4-dichlorophenoxyacetic acid (2,4-D) in rat and dog. Xenobiotica. 33:805–821.

Vardia HK, Rao S, Durve VS. 1984. Sensitivity of toad larvae to 2,4-D and endosulfan pesticides. Archiv Fur Hydrobiologiey. 3:395–400.

Vonier PM, Crain DA, McLachlan JA, Guillette LJ Jr. Arnold SF. 1996. Interaction of environmental chemicals with the estrogen and progesterone receptors from the oviduct of the American alligator. Environ. Health Persp. 104:1318–1322.

Wolf CJ, Hotchkiss A, Ostby JS, LeBlanc GA, Gray LE Jr. 2002. Effects of prenatal testosterone propionate on the sexual development of male and female rats: a dose-response study. Toxicol Sci. 65:71–86.

Xie L, Thrippleton K, Irwin MA, Siemering GS, Mekebri A, Crane D, Berry K, Schlenk D. 2005. Evaluation of estrogenic activities of aquatic herbicides and surfactants using an rainbow trout vitellogenin assay. Toxicol Sci. 87:391–398.

Supplementary Appendices

Supplementary Appendix I SEARCH TERMS

CAS numbers searched included: (94-75-7 or 1929-73-3 or 94-80-4 or 2008-39-1 or 5742-19-8 or 1928-43-4 or 1713-15-1 or 25168-26-7 or 94-11-1); 5742-17-6 or 2702-72-9 or 18584-79-7 or 2569-01-9); (5742-17-6 or 2702-72-9 or 18584-79-7 or 2569-01-9); and chemical names:(2(w)4(w)('d' or dichlorophenoxyacet? or dichlorophenoxy acet? or dichloro phenoxy acet? or dichloro phenoxyacet?))

The following search terms were used to identify endocrine relevant papers:

S (ESTROGEN? OR OESTROGEN? OR ESTRADIOL OR OESTRADIOL OR ESTROGEN RECEPTORS+NT/CT OR OESTROGEN RECEPTORS+NT/CT OR OESTROGENS+NT/CT OR ESTROGENS+NT/CT OR HELA 9903 OR HELA9903)

S AND (ANTIESTROGEN? OR ANTIOESTROGEN?)

S AND (ANDROGEN? OR TESTOSTERONE OR ANTIANDROGEN? OR ANTITESTOSTERONE)

S AND (H295R OR STEROIDOGEN? OR STEROID(1W)(SYNTHESIS OR PATHWAY OR METABOLI?)

OR HORMONE(1A)(SYNTHESIS OR PATHWAY OR METAB? OR BIOSYNTH?))

S AND (AROMATASE OR CYTOCHROME(W)(P450 OR P 450) OR CYP19)

S ((UTERINE OR UTERUS OR FEMALE GENITALIA OR GENITALIA, FEMALE+NT/CT OR

FEMALE GENITAL SYSTEM+NT/CT) AND (SIZE OR WEIGHT OR HYPERTROPH? OR ENLARG?))

S AND HERSHBERGER

S AND (PROSTATE OR SEMINAL VESICLE OR LEVATOR(1A)BULBOCAVERN?

OR COWPERS GLAND OR PENIS)

S AND (TESTIS OR TESTES OR EPIDIDYM? OR MALE GENITALIA OR GENITALIA, MALE+NT/CT) AND

(SIZE OR WEIGHT OR HYPERTROPH? OR ENLARG? OR HYPOTROPH? OR SHRINK?)

S MALE REPRODUCTIVE SYSTEM+NT/CT OR COWPERS GLAND OR LEVATOR(1A)BULBOCAVERN?

OR PENIS OR GENITALIA, MALE+NT/CT OR MALE GENITALIA SYSTEM+NT/CT

S AND (RAT OR RATTUS) AND (PUBERTY OR PUBESC? OR SEX? MATUR?

OR PRE-PUTIAL OR VAGINA?(1A) OPEN?)

S AND (THYROID OR TSH OR THYROTROPIN OR THYROXIN? OR TRIIODOTHYRONINE

OR THYROID HORMONES+NT/CT

S AND (CHOLINESTERASE OR CHOLINE ESTERASE)(1A)(INHIBIT? OR BLOCK? OR ANTAGONI?)

OR 9001-08-5(S)(BLOCK? OR INHIBIT? OR ANTAGONI?)

S (FROG OR AMPHIB? OR ZENOPUS OR RANA OR RANIDAE OR ANURA)

S (TOAD OR TADPOLE OR METAMORPHOSIS OR ANURA+NT/CT OR TOAD+NT/CT OR TADPOLE)

S METAMORPHOSIS

S AND (MINNOW OR PIMEPHALES OR MEDAKA OR ORYZIAS OR

CYPRINIDON OR POECILIA OR GUPPY OR ZEBRAFISH)

S(ZEBRA(W)(FISH OR DANIO) OR (DANIO OR BRACHIDANIO OR CYPRINUS)(W) RERIO)

S PITUIT? OR HYPOTHAL? OR HPG AXIS OR HYPOTHAL? PITUITARY GONADAL AXIS

OR HYPOPHYSIS OR HYPOPHYSEAL OR PITUITARY GLAND+NT/CT

OR HYPOTHALAMUS+NT/CT OR PITUITARY DISEASES+NT/CT

OR HYPOTHALMIC DISEASES+NT/CT OR HYPOTHALAMO-HYPOPHYSEAL SYSTEM+NT/CT

Supplementary Appendix II: *In vitro* Studies with Klimisch Scores of 1 or 2[1]

A. Regulatory Toxicology (EDSP Tier 1) *in Vitro* Studies

As discussed previously in the publication, the high concentration tested in the EDSP Tier I *in vitro* assays (except steroidogenesis) was restricted below the maximum concentration specified in the EPA Guidelines, such that the concentration did not exceed the TSRC established through TK in the *in vivo* mammalian EOGRT Study.

LeBaron, *et al.*, 2011a; (published in Coady *et al.*, 2014)

In an estrogen receptor (ER) binding assay following the US EPA Guideline (US EPA, 2009a) with the exception regarding the high concentration noted above, uterine cytosol from Sprague Dawley rats was used as the source of ER to conduct a saturation binding experiment and a competitive binding experiment. The saturation binding experiment was conducted to demonstrate that the ER was present in reasonable numbers and was functioning with appropriate affinity for the radiolabeled ligand (^{3}H-17β-estradiol). Subsequently, the competitive binding experiment was conducted in three independent assay runs to measure the binding of a single concentration of [^{3}H]-17β-estradiol (1 nM) in the presence of increasing concentrations of 2,4-D (98.5% a.i) ranging from 10^{-11} M to 10^{-4} M. Ethanol was used as a vehicle at a final concentration of <3% (v/v). The adequacy of the experimental conditions for the detection of ER binding was confirmed through the testing of reference chemicals: 17β-estradiol (strong positive control, radioinert), 19-norethindrone (weak positive control), and octyltriethoxysilane (negative control). In addition, control tubes were treated with the solvent used to dissolve the test material (*i.e.*, ethanol) for determination of full binding capacity and calculation of relative binding activity.

In the saturation binding experiment, the maximum binding capacity (B_{max}) was 59.28 fmol/100 µg protein and the dissociation constant (K_d) was 0.1032 nM. These results were within the acceptable range from the validation studies. All other saturation binding performance criteria and results from the competitive binding experiment indicated acceptable performance of the assay. The Scatchard plot indicated a linear response across the concentrations of ligand added. Nonspecific binding as a percent of total binding was 1.7%-8.6% across the entire concentration range in the saturation binding experiment.

In the competitive binding experiment, there were no appreciable alterations in 17β-estradiol ER binding activity at 2,4-D concentrations ranging from 10^{-11} M to 10^{-4} M in three independent runs of the assay. Summary results for the ER binding assay of 2,4-D are shown in **Table S1**. No estimated log IC_{50} and relative binding affinity (RBA) were calculated for 2,4-D because of the absence of binding activity. The logIC_{50} values for 17β-estradiol and 19-norethindrone were -9.0 and -5.5, respectively. Compared to 17β-estradiol, the RBA for 19-norethindrone was 0.034%. The reference chemicals (17β-estradiol, 19-

[1] See Publication **Tables 3** and **4** for Klimisch scores for the regulatory EDSP and published studies, respectively.

norethinedrone, and octyltriethoxysilane) met the QC performance criteria established in the test guideline in all instances, except for some slight deviations for 19-norethinedrone in Run 3. However, because the acceptance ranges for the weak positive control are based on use of norethynodrel, these slight deviations from the suggested ranges are considered minor.

Based on the results of this assay, 2,4-D was classified as negative for ER binding.

LeBaron and Kan, 2011; Coady *et al.*, 2014

In an estrogen transcriptional activation assay complying with the EPA Guideline (US EPA, 2009b) with the exception regarding the high concentration noted above, human ERα-HeLa-9903 cells cultured *in vitro* were exposed to 2,4-D (98.5% a.i.) at concentrations of 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5} and 10^{-4} M in DMSO (0.1% (*v/v*)) for 24 ± 2 hours. The experiments were performed using 96-well plates and each 2,4-D concentration was tested in triplicate. Cells were exposed to the test agent for 24 ± 2 hours to induce reporter (luciferase) gene products. Luciferase expression in response to activation of the estrogen receptor by 2,4-D was measured upon addition of a luciferase substrate and detection with a luminometer with acceptable sensitivity.

2,4-D was tested in four independent runs of the ER transcriptional activation assay. Cytotoxicity and precipitation were not observed at any of the test concentrations in these four runs.

The adequacy of the experimental conditions for the detection of ER agonism was confirmed using the following reference chemicals: 17β-estradiol (strong agonist), 17α-estradiol (weak agonist), 17α-methyltestosterone (very weak agonist), and corticosterone (negative control). In addition, vehicle control cultures were treated with the solvent used to dissolve the test material (*i.e.*, DMSO) for determination of basal transcriptional activity. Results of the four concurrent reference chemicals included in each experiment generally fell within the acceptable ranges. There was slight variability between assays, but the overall robustness of the responses for 17β-estradiol, 17α-estradiol, 17α-methyltestosterone, and corticosterone indicated that each assay run included in this assessment performed as expected. In most cases where values were outside of defined ranges, the values indicated increased sensitivity of the assay system (*i.e.*, decreased log PC_{50} and log PC_{10} for 17α-methyltestosterone and 17α-estradiol).

The RPC_{max} values for 2,4-D (maximum level of response induced by the test chemical expressed as a percentage of the response induced by 1 nM 17β-estradiol on the same plate) were 0%, 8.8%, 5.3%, and 7.1% in Runs 1, 2, 3 and 4, respectively. The mean RPC_{max} of the four assays was 5.3%. At concentrations of 2,4-D ranging from 10^{-10} M to 10^{-4} M, there were no appreciable increases in estrogen receptor-mediated transcriptional activity. Based on these results, 2,4-D was judged to be negative for estrogen receptor-mediated agonism.

LeBaron *et al.*, 2011b; (published in Coady *et al.*, 2014)

In an androgen receptor (AR) binding assay following EPA Guidelines (US EPA, 2009c) with the exception regarding the high concentration noted above, ventral prostate cytosol from Sprague-Dawley rats was

used as the source of AR to conduct a saturation binding experiment and a competitive binding experiment. Saturation binding experiments were conducted to demonstrate that the AR was present in reasonable numbers and was functioning with appropriate affinity for the radiolabeled ligand (^{3}H-R1881). Subsequently, the competitive binding experiment was conducted in three independent assay runs to measure the binding of a single concentration of ^{3}H-1881 (1 nM) in the presence of increasing concentrations of 2,4-D (98.5% a.i.) ranging from 10^{-11} M to 10^{-4} M. Ethanol was used as a vehicle at a final concentration of <3% (v/v). The adequacy of the experimental conditions for the detection of AR binding was confirmed through the testing of reference chemicals: R1881 (strong positive control, radioinert) and dexamethasone (weak positive control), both of which met established quality control criteria. In addition, control tubes were treated with the solvent used to dissolve the test material (*i.e.,* ethanol) for determination of full binding capacity and calculation of relative binding activity.

In the saturation binding experiment, the maximum binding capacity (B_{max}) was 3.245 fmol/100 µg protein and the dissociation constant (K_d) was 0.4641 nM. Although these values were slightly below the range of values from the validation studies, the results were highly reproducible and all other performance criteria indicated acceptable performance of the assay. The Scatchard plot indicated a linear response across the concentrations of ligand added. Nonspecific binding as a percent of total binding was less than 20% across the entire concentration range in the saturation binding assays (range: 6.2-19.8%), with one exception (24.6%) at the high concentration (10 nM) in one assay.

In the competitive binding experiment, there were no appreciable alterations in R1881 AR binding activity at 2,4-D concentrations ranging from 10^{-11} to 10^{-4} M in three independent runs of the assay. Summary results for the AR binding assay of 2,4-D are shown in **Table S2**. The estimated log IC_{50} and relative binding affinity (RBA) for 2,4-D was not calculated due to the absence of activity. The log IC_{50} values for R-1881 and dexamethasone were -9.0 and -4.4, respectively. Compared to R1881, the RBA for dexamethasone was 0.0027 %. In all instances, R1881 and dexamethasone met the QC performance criteria established in the test guideline.

Based on the results of this assay, 2,4-D was classified as negative for AR binding.

Coady and Sosinski, 2011; (Published in Coady *et al.*, 2014)

In an aromatase assay conducted according to US EPA Guidelines (US EPA, 2009d) with the exception regarding the high concentration noted above, 2,4-D (98.5% a.i) was incubated with human recombinant aromatase and tritiated androstenedione (1-β[^{3}H(N)]-Androst-4-ene-3,17-dione; [^{3}H]ASDN) in ethanol at concentrations of 0, 10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, $10^{-4.5}$ or 10^{-4} M for 15 minutes to assess the effect of 2,4-D on aromatase activity. Aromatase activity was determined by measuring the amount of tritiated water produced at the end of 15 minute incubations for each concentration of chemical. Tritiated water was quantified using liquid scintillation counting (LSC). Three independent runs were conducted and each run included a full activity control, a background activity control, a positive control series (10^{-10} - 10^{-5} M) using 4-hydroxyandrostenedione (4-OH ASDN), a known inhibitor of aromatase, and the test chemical series (10^{-10}- 10^{-4} M) with 3 replicates per concentration. Standard curves generated from Runs 1, 2 and 3 with the known inhibitor, 4-OH ASDN, generally met performance criteria as described in the US EPA guideline. In addition, the standard curves generated for each run had a goodness of fit values that were equal to or exceeded 99%. The full activity and background controls in each of the runs were within the recommended ranges for the assay. Thus, Runs 1, 2, and 3 of the aromatase assay with 2,4-D were considered acceptable for determining whether or not the test material has the potential to inhibit aromatase activity *in vitro.*

The average response from three independent runs of the human recombinant aromatase assay with 2,4-D did not fit the four parameter regression model. Additionally, average aromatase activity for 2,4-D was similar to that of the full activity controls at all concentrations tested. Based on these results, 2,4-D was classified as a non-inhibitor of aromatase activity.

LeBaron *et al.*, 2011c; (Published in Coady *et al.*, 2014)

In a steroidogenesis assay following US EPA Guidelines (US EPA, 2009e), H295R cells cultured *in vitro* in 24-well plates were incubated with 2,4-D (98.5% a.i.) at concentrations of 0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, and 100.0 µM in triplicate for 48 hours. The test chemical vehicle was DMSO (0.1% (v/v)). The highest 2,4-D concentration tested is the assay limit concentration of 10^{-4} M. Testosterone and estradiol levels were measured using LC/APPI-MS/MS. Positive control chemicals prochloraz (an inhibitor) and forskolin (an inducer) at two concentration levels were tested with each run of the assay. In addition, control cultures were treated with the solvent vehicle for determination of acceptable basal hormone production levels (*i.e.*, minimally ~500 pg/ml testosterone and ~20 pg/ml estradiol). The test chemical, reference chemicals, and solvent controls were tested in replicates of 3/plate.

Three independent runs of the steroidogenesis assay were performed with 2,4-D. No cytotoxicity > 20% or precipitation were observed at any of the test concentrations of 2,4-D. The steroidogenesis assay of 2,4-D was conducted in accordance with OPPT 890.1550 test guidelines and EPA GLP regulations. The

assay performance criteria were met for all three runs of the assay, with minor exceptions not considered to affect the interpretation of the study results.

Measured concentrations of testosterone and estradiol in the culture media following 2,4-D exposure in Runs 1-3 are shown in **Table S3**. At concentrations of 2,4-D ranging from 10^{-10} M to 10^{-4} M, testosterone production was not statistically different than that of the solvent controls. The limit concentration of 2,4-D (100 µM, 10^{-4} M) resulted in a small, statistically significant increase in estradiol production that was reproducible across assays.

Based on these data, 2,4-D was judged to produce slightly increased estradiol production at the highest (limit) concentration evaluated and not at any of the lower concentrations. This slight (1.2-fold) increase in estradiol production, however, did not meet the cut-off criteria for increase (*i.e.*, 1.5-fold) of estradiol established in the EDSP validation program as indicating a positive response for the steroidogenesis assay (Hecker et al., 2008) and is not considered biologically relevant.

B. *In Vitro* Studies Identified in the Published Literature

Kojima *et al.*, 2004

Two hundred pesticides, including 2,4-D (purity between 95-100%), were tested for agonist and antagonist activity at the human ERα, human ERß, and human AR. Chinese hamster ovary cells (CHO K1) were plated in 96-well plates at a density of 8,400 cells per well in DMEM/F-12 media containing 5% FBS pretreated with dextran-coated charcoal. One day later, the cells were transiently transfected with ERα, ERß, or AR. Cells were simultaneously transfected with a luciferase reporter plasmid containing either an estrogen responsive element (ERE) or an androgen responsive element (ARE) upstream of the firefly luciferase gene, and a plasmid containing the *Renilla* luciferase gene (used as an internal control for transfection efficiency). Three hours post-transfection, cells were treated with 10^{-8} – 10^{-5} M of each test compound or 0.1% DMSO (vehicle control). To evaluate antagonist activity at the ERα, ERß, and AR, test compounds were added in the presence of either 10^{-11} M E2, 10^{-10} M E2, or 10^{-10} M DHT, respectively. After 24 hours incubation, the cells were lysed and firefly luciferase activity measured and normalized against expression of *Renilla* luciferase activity. A test compound was determined to have agonist activity if it showed an activity level within the range of concentrations tested that was ≥20% that induced by 10^{-10} M E2, 10^{-9} M E2, or 10^{-9} M DHT at the ERα, ERß, or AR, respectively. Similarly, a test compound was determined to have antagonist activity if it inhibited the activity of the agonist by at least 20% within the range of concentrations tested. Mean results of three independent experiments were reported. Within the concentration range of 10^{-8} – 10^{-5} M, 2,4-D did not demonstrate agonist or antagonist activity at the ERα, ERß, or AR.

This study is considered to be of high quality and demonstrated an absence of ERα transactivation by 2,4-D, a lack of antagonist activity at the ERα and the absence of both agonist and antagonist activity at ERß, and an absence of ability to transactivate the AR or antagonize the activity of DHT at the AR.

Lin and Garry, 2000

Sixteen agrochemicals, including 2,4-D and 2,4-D isooctyl ester, were assessed for the ability to induce proliferation of an estrogen-responsive human breast cancer cell line. To examine cytotoxicity, MCF-7 cells were seeded in DMEM/F-12 medium (without phenol red; supplemented with 10% FBS [not charcoal-dextran treated] and 1% gentamicin) at a density of 10^5 cells/ml in 24-well plates. Test chemicals or solvent alone were added within 2-4 hours of incubation. After 72 hours in culture, the cells were evaluated for cell number and viability by flow cytometry. Results of cytotoxicity assays were not shown but concentrations tested reported to be non-cytotoxic. To assess cell proliferation, 10^4 MCF-7 cells/ml were seeded into 24 well plates using medium containing either 10% charcoal-dextran treated FBS or 10% non-charcoal-treated FBS. Forty-eight hours later, cells were treated with 10^{-9} M E2, solvent vehicle, or dilutions of test chemicals (0.1 to 10 µg/mL; duplicate wells per sample) and cultured for an additional seven days. The final volume of solvent was <0.1% of total volume. At end of culturing period, cells were harvested and examined for cell number and viability by flow cytometry. In all cases, three separate experiments were run with at least duplicate samples per concentration tested. Commercial grade 2,4-D LV4 (66.2% purity 2,4-D isooctyl ester) and 2,4-D amine (46.5% purity 2,4-D dimethylamine) both induced MCF-7 cell proliferation, with maximum proliferation at concentrations of 1 µg/ml. Reagent grade 2,4-D isooctyl ester and reagent grade 2,4-D acid, however, had no effect on MCF-7 cell proliferation, suggesting that the proliferative effect was due to additives or contaminants present in the commercial products.

Sun *et al.*, 2012

Sun *et al.* (2012) developed luciferase reporter gene assays to measure the effects of 2,4-D on estrogen receptor alpha (ERα), androgen receptor (AR) and thyroid hormone receptor (TR) activities. Vero cells, derived from African green monkey kidney epithelium, were used because they do not express any of these receptors endogenously. Cells were transfected with one of the ERα, AR, or TR luciferase reporter plasmids and subsequently treated with 0.1-, 1-, 10- and 100-fold concentrations of 2,4-D maximum contamination level (MCL) of 0.03 mg/L in Chinese drinking water, i.e., cells were exposed to 0.003, 0.03, 0.3, and 3.0 mg/L 2,4-D. Agonist activity was measured as 20% relative effective concentration, which was defined as the concentration at which 2,4-D showed 20% of the maximum activity of E2, testosterone, or T3. Similarly, antagonist activity was measured as 20% relative inhibitory concentration when co-treated with $1x10^{-6}$ mg/L E2, $5x10^{-5}$ testosterone, or $5x10^{-4}$ mg/L T3.

2,4-D exhibited no agonist or antagonist activity for ERα even at concentrations 100 times the MCL. There was also no detectable androgenic or anti-androgenic activity at any of the 2,4-D concentrations tested; however, at 100x the MCL, 2,4-D was reported to enhance the effects of testosterone in the AR antagonist assay. The health significance of this finding is uncertain and it was limited only to the highest concentration tested, and in an assay system designed to test anti-androgenicity and not androgenicity. The concentration showing this effect was roughly equivalent to a 2,4-D serum concentration resulting from an *in vivo* dose greater than 5 mg/kg/day (Saghir *et al.*, 2013), the overall

2,4-D NOEL dose used as the point of departure dose to set the chronic RfD for 2,4-D, and thus this finding is not relevant to human health risk assessment.

Supplementary Appendix III

Ecotoxicological Studies with Klimisch Scores of 1 or 2[2]

A. Regulatory Toxicology (EDSP Tier 1) Studies

Coady *et al.*, 2010; (published in Coady *et al.*, 2013)

The Tier I EDSP Amphibian Metamorphosis Assay (AMA) is an assay in frogs designed to focus on potential thyroid effects. The study design corresponded with Guidelines: OPPTS 890.1100 (US EPA 2009f) and OECD 231. In brief, African clawed frog (*Xenopus laevis*) tadpoles were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days. Nominal test concentrations were selected based on prior acute toxicity studies with *Xenopus laevis* embryos and *Rana pipiens* larvae, which showed that the required high concentration equivalent to 1/3 of the LC50 (estimated maximum tolerated concentration (MTC)) was close to the 100 mg/L guideline limit concentration. The negative control was untreated laboratory dilution water. The study design is summarized in **Table S4**.

The concentration of 2,4-D in the exposure solutions was measured from all replicate tanks on a weekly basis during the test. Mean measured concentrations were 0.273, 3.24, 38.0 and 113 mg acid equivalents (a.e.)/L, representing 68.3, 81.0, 95.0 and 113% of nominal concentrations, respectively. Test substance was not detected in the negative control (Limit Of Quantitation (LOQ) = 0.120 mg a.e./L). A decline in measured concentrations was most likely due to biodegradation, which reduced 2,4-D concentrations in all exposure groups except for the high concentration, and proportionately had a greater effect at the two lowest nominal concentrations.

Hind limb length was normalized by snout-vent length to account for the effects of growth. Length and weight data for tadpoles reaching stages greater than Nieuwkoop and Faber (NF) stage 60 were excluded from statistical analyses in accordance with guideline requirements due to the typical drastic morphological changes in these parameters due to metamorphic progression at this stage and above (US EPA 2009f; Nieuwkoop and Faber, 1994).

The only performance criterion that was not met consistently in the AMA with 2,4-D was that a total of six replicate test vessels (4 replicate test vessels in the 0.273 mg./L 2,4-D treatment group and 2 replicate test vessels in the 3.24 mg /L 2,4-D treatment group) had measured concentrations of 2,4-D with coefficients of variation that exceeded 20%. This was likely due to biodegradation of the test material in the test vessels. Since the validity criteria and most of the performance criteria were fulfilled, and test concentrations were characterized, the study is considered valid, of high quality (Klimisch score of 1), and useful for assessing potential thyroid effects in the WoE.

There were no exposure related effects on survival or behavior. There were no statistically significant effects of 2,4-D on wet weight, snout-vent length (SVL), normalized hind limb lengths (HLL) or median NF developmental stage at either day 7 or day 21. There were no exposure-related histopathologic changes

[2] See Publication **Table 7** for Klimisch scores and comments for regulatory and published ecotoxicological studies

in the thyroid gland in any of the dosed groups, including no evidence of glandular atrophy, hypertrophy or follicular cell hyperplasia. The incidence of tall columnar cells lining the follicles (follicular cell hypertrophy) did not show any exposure-related differences and was interpreted to be within normal limits at all concentrations of 2,4-D. There was no evidence of any inflammatory or degenerative changes in the thyroid glands examined in any treatment group.

All other histopathologic criteria of all 2,4-D exposed tadpoles, including the overall size of the gland, the follicular lumen area, amount and type of colloid, and the follicular cell type and arrangement, were comparable to those of the controls.

In sum, there were no biologically significant exposure-related effects on the thyroid gland or other endpoints of this assay that are potentially subject to perturbations in the Hypothalamus-Pituitary-Thyroid (HPT) axis (*i.e*,.normalized hind limb length developmental stage, occurrence of asynchronous development). Thus, there is no evidence of a potential interaction with the HPT axis in this Tier 1 EDSP AMA tested to the limit concentration of 100 mg 2,4-D/L.

Marino *et al.*, 2010 (published in Coady *et al.*, 2013)

Marino *et al.* (2010) tested 2,4-D in a Fish Short Term Reproduction Assay (FSTR assay) (US EPA 2009g). Details of this study, which is designed to assess potential interactions with the estrogen, androgen or steroidogenesis pathways, are provided in Coady *et al.*, 2013. The study was conducted in compliance with OPPTS 890.1350 and OECD 229. Sexually mature adult fathead minnows (*Pimephales promelas*) were exposed to 2,4-D (98.6% purity) under continuous flow-through conditions for 21 days. Nominal test concentrations were selected based on acute toxicity tests and an early life stage test with fathead minnows (Alexander *et al.*, 1983; Mayes *et al.* 1990). The negative control was untreated laboratory dilution water. The study design is summarized in **Table S5**. Survival, fecundity, fertilization success, and general observations of health were noted daily during both the 14-day pre-exposure period and the 21-day exposure period. At termination on day 21, lengths and weights of surviving fish were measured, as well as secondary sex characteristics (including tubercle scores), gonado-somatic index (GSI), histopathology of gonads, and plasma vitellogenin concentrations.

The concentration of 2,4-D in the exposure solutions was measured from all replicate test vessels at weekly intervals during the test using a validated analytical method. Mean measured concentrations were 0.245, 3.14, 34.0 and 96.5 mg a.e./L, representing 61.3, 78.5, 85.0 and 96.5% of nominal concentrations, respectively. Test substance was not detected in the negative control (LOQ = 0.10 mg a.e./L). A decline in measured concentrations was most likely due to biodegradation, which reduced 2,4-D concentrations in all treatment groups but proportionately had a greater effect at the two lowest nominal concentrations.

The only performance criterion that was not met consistently was that a total of five replicate test vessels had measured concentrations of 2,4-D with coefficients of variation that exceeded 20%, likely due to degradation as described above. Since the validity criteria and most of the performance criteria were fulfilled, the study is considered valid, of high quality (Klimisch score of 1), and useful for assessing potential estrogen or androgen pathway interactions or effects on steroidogenesis in the WoE.

The overall results of the study are summarized in **Table 5** in the publication. The only significant effect compared to the controls was a decrease in fecundity (considered a non-specific finding) among fish exposed to 96.5 mg a.e./L 2,4-D. In the absence of effects upon the other, more specific endocrine-mediated endpoints, the isolated effect on fecundity at 100 mg a.i./L is most likely due to systemic toxicity and a generalized stress response. This concentration is relatively high (approximately 1/3 of the acute LC50 value in fish), is the limit concentration for the FSTR assay, and is a concentration which exceeds the Maximum Acceptable Toxicant Concentration (MATC) for larval fish survival in an early life stage toxicity test with fathead minnows (Mayes *et al.*, 1990).

In conclusion, 2,4-D does not appear to interact with the estrogen, androgen or steroidogenic pathways, or with the HPG axis, in fathead minnows tested up to the limit concentration in this EDSP Tier 1 FSTR assay.

Mitchell *et al.*, 1999

An avian single generation reproductive toxicity study (Mitchell *et al.*, 1999) of 2,4-D shows limited toxicity to quail and a lack of potentially endocrine-related effects.

In this study at Wildlife international, 2,4-D Acid (96.9% pure) was administered to adult Northern Bobwhite quail (*Colinus virginianus*) for 21 weeks via the diet at the following nominal concentrations: 0, 160, 400, and 1000 ppm active ingredient. Analysis of feed during the study indicated that measured concentrations of 2,4-D were close to nominal values (92-96% nominal). All adult birds were observed daily for signs of toxicity and abnormal behavior. Adult body weights were measured at test initiation, weeks 2, 4, 6, 8, and at adult termination. Feed consumption was measured weekly throughout the test. Following egg production, eggs were set weekly for incubation. During incubation, a subset of the eggs was assessed weekly for eggshell thickness. Eggs were additionally examined to detect eggs cracks, abnormal eggs, infertile eggs or embryo mortality. Eggs were hatched on Day 21 of incubation. Hatchlings were assessed (by pen) for group body weight. Weight of surviving offspring was again assessed at 14 days of age.

Study findings are summarized in **Table 6** of the Publication. There were no treatment-related mortalities, overt signs of toxicity or treatment-related effects upon body weight or feed consumption at any of the concentrations tested. Additionally, there were no treatment-related effects upon any of the reproductive parameters measured at the 160, 400 or 1000 ppm a.i. test concentrations, including fertility, egg shell thickness, embryo-mortality or hatching success. The no-observed effect concentration for northern bobwhite exposed to 2,4-D acid in the diet during the study was 1000 ppm a.i., the highest concentration tested. This study is considered valid; it predicts a very low hazard of reproductive toxicity of 2,4-D to birds and a low likelihood of endocrine-related effects on birds.

B. Other Ecotoxicological Studies Identified in the Published Literature

Reptiles

Crain *et al.*, 1997

American alligator (*Alligator mississippiensis*) eggs collected from Lake Woodruff, Florida (relatively pristine environment) were topically dosed with 17β-estradiol (positive control) or the herbicides atrazine or 2,4-dichloro-phenoxyacetic acid (2,4-D) (Chem Service, West Chester, PA; 97.6% purity) prior to the time of sexual differentiation. Atrazine will not be discussed further. Each egg received a single dose of 0.014, 0.14, 1.4, or 14 ppm estradiol or 0.14, 1.4, or 14 ppm 2,4-D applied topically to the eggshell in 50 ml of 95% ethanol. (This is a technique frequently used to transport compounds inside reptilian eggshells.) An ethanol solvent control and an untreated control were also included in the study design. Upon pipping, chorio-allantoic fluid (CAF) was collected and analyzed for estradiol and testosterone using radioimmunoassay. Hatchlings were then housed individually for 10 days. Just prior to euthanasia, blood was collected for estradiol and testosterone analyses. Following euthanasia, the right gonadal-adrenal mesonephros (GAM) complex was removed for an aromatase bioassay. The left GAM was removed and analyzed histopathologically to determine sex.

Results indicated that *in* ovo exposure of American alligators to 2,4-D did not alter sexual differentiation and had no effect on plasma steroids, or on gonadal aromatase activity. This study is included in the WoE because hormone levels were directly measured post 2,4-D exposure, the method of application has been used previously and penetration of test material demonstrated, and a positive control was tested.

Crain *et al.*, 1999

This study extends from the study by Crain *et al.*, (1997) above. The same test organisms, methods, and test compounds were used; however, hepatic aromatase activity was measured in hatchling alligators and a more complete histopathological analysis of the gonad was completed. Results indicated that 2,4-D exposure *in ovo* prior to the time of sexual differentiation had no effect on hepatic aromatase activity or testicular histopathology. This study is included in the WoE because aromatase was directly measured post 2,4-D exposure.

Spiteri *et al.*, 1999

This study extends from the study by Crain *et al.*, (1997) and Crain *et al.*, (1999) above. The same test organisms, methods, and test compounds were used. Five eggs from each dose-treatment group were incubated at either a temperature that produces 100% males (33°C) or a temperature that produces 100% females (30°C). At ten days post-hatch, gonadal histology and hepatic steroidogenic activity were assessed in the hatchlings. Hatchlings in all the estradiol treatment groups were 100% female regardless of their incubation temperature. Hatchlings in the 2,4-D treatment group developed as males or females, as was expected based on their incubation temperature. Thus, there were no signs of sex reversal due to herbicide exposure. Compared to controls, 14 ppm estradiol caused a significant increase in the degeneration of the ovarian medulla and a significant increase in the height of Müllerian

duct epithelial cells. (The high dose of estradiol necessary to elicit these responses indicates that these histopathological endpoints are not as sensitive as gender reversal.) No significant histopathological differences were noted among the gonads of hatchlings exposed to 2,4-D. The publication also reiterated the results from hepatic aromatase assays of the hatchling alligators summarized in Crain *et al.*, 1999 above; it is not clear that any new data were developed on this parameter. The design and methods employed in the present study indicate that the study is scientifically valid. The results of this study suggest that egg exposure to concentrations up to 14 ppm 2,4-D do not cause significant alterations in gonadal structure or hepatic steroidogenic enzyme activity of hatchling American alligators. In contrast, all doses of estradiol (even as low as 0.014 ppm) caused feminization of prospective males. These data have relevance for the weight of evidence for endocrine activity since steroidogenesis and gonadal histopathology are relevant endpoints; also the test system was demonstrated to respond to a positive control. Under the conditions of exposure, 2,4-D, did not affect steroidogenesis or ovarian activity.

Supplementary Appendix IV

Mammalian Toxicological Studies with Klimisch Scores of 1 or 2[3]

A. Regulatory Toxicology Studies

Reproductive Toxicity Evaluations

Marty *et al.*, 2010; Marty *et al.*, 2013

The EOGRT study of 2,4-D (Marty *et al.*, 2010, published in Marty *et al.*, 2013) was specifically designed to provide sufficient information to assess whether endocrine targets are, in fact, altered with *in vivo* exposure, and to provide the basis for robust risk assessment of 2,4-D, including risk assessment protective for any potential endocrine effects. The study design was based on a modified reproductive toxicity study protocol that evaluated multiple endpoints across life stages, including estrogen, androgen and thyroid effects (Cooper *et al.*, 2006). This study design provides a reliable basis for establishing the potential of 2,4-D to interact with the estrogen, androgen or thyroid pathways and is considered a Tier 2 EDSP-equivalent assessment.

As noted, this study incorporated the extensive toxicokinetic information on 2,4-D in dose setting, to inform dose selection. The high dose dietary exposure concentrations, which differed for males and females, were predicted to just exceed the TSRC. This study also included additional measures of internal dosimetry. Based on these latter data, the high dose in males (800 ppm or approximately 40 mg/kg/day) adequately approached the TSRC, but the high dose in females (600 ppm or approximately 25-30 mg/kg/day) clearly exceeded the TSRC.

This study is reported in depth in a publication by Marty *et al.*, 2013. A summary of the study design and endocrine-related parameters for the EOGRT study follows:

CD®rats (27/sex/dose) were fed diets containing 0, 100, 300, and 600 (females) or 800 (males) ppm 2,4-D (98.75% purity), supplying approximately 0, 7, 21, or 40 mg/kg/day 2,4-D for adult females and 0, 6, 17, or 45 mg/kg/day 2,4-D for adult males for four weeks prior to breeding and continuing through breeding (up to two weeks), gestation (three weeks), and lactation (three weeks). Exposure of P (parental generation) males continued to ensure dosing covered a full spermatogenic cycle. P females were exposed until LD 22 (the end of the lactation period). P males and females were evaluated for systemic toxicity, as well as functional and structural evaluations of the reproductive systems. In addition, a satellite group of P females (12/dose) was exposed during pre-breeding, breeding, and GD 0-17, when they were euthanized to assess gross pathology, clinical pathology, thyroid hormones, TK and selected reproductive parameters during gestation.

[3] Klimisch scores and comments for regulatory developmental, subchronic and chronic toxicity studies are in publication **Table 10**, and for published mammalian toxicological studies in publication **Table 11**.

Table S6 summarizes the initial group assignments:

Dietary concentrations of 2,4-D were adjusted during lactation (LD 7-21) by a factor of 2x or 3x to account for the large increase in feed consumption (2-3-fold) typical for rat dams in mid- to late lactation- LD 7-14 (Hanley *et al.*, 1985). Diets to weanlings were also adjusted until PND 35. These dietary concentration adjustments were intended to maintain more consistent doses of 2,4-D (mg/kg/day basis) across study phases and were based on TK information from the TK/rangefinding study of 2,4-D (Saghir *et al.*, 2008a). Dose adjustments are shown in **Table S7.**

Selected F1 offspring were maintained on the test diet until PND 60 (10/sex/dose), ~PND 70 (Sets 1a; 10/sex/dose and 2a; 10/sex/dose) or ~PND 90 (Set 2b; 10/sex/dose) and ~PND 139 (Set 3; 20/sex/dose). F1 Offspring were evaluated for effects on systemic toxicity (Set 1a), the nervous system (Set 1b), immune system (Set 2a and 2b), and the reproductive system and thyroid function (Set 3). 2,4-D TK was assessed in the Set 3 F1 offspring on PND 63 and 84.

Systemic toxicity: Systemic toxicity was limited to body weight decreases and decreased feed consumption in P females during lactation (evident before diet concentration adjustment), and decreased pup body weights in pups at 600 ppm. Animals during these life stages still received more compound on a mg/kg/day basis (despite dose adjustment) than did animals in other life stages and high-dose females were considered to have exceeded the TSRC. Renal toxicity was observed at the high dose in males and females, consistent with other studies of 2,4-D. Very slight renal effects were seen at the mid-dose; these are considered toxicologically insignificant. There was no indication of developmental neurotoxicity or developmental immunotoxicity in the study.

Reproductive toxicity: 2,4-D had no effects on estrous cyclicity or P reproductive indices, including mating, fertility, time to mating, gestation length, pre- and post-implantation loss and *corpora lutea* number (examined in satellite dams). There were slight non-statistically and non-biologically significant decreases in the fertility indices at 300 and 600 ppm in the P animals; these differences were within the laboratory historical control data (HCD) range. Litter size, pup survival, female reproductive organ histopathology and ovarian follicle counts were unaffected by 2,4-D exposure.

Decreased bilateral testis size was found in one P male in each of the 300 and 800-ppm groups. This finding was not considered exposure-related; the incidence was well within the laboratory HCD range. A similar but unilateral finding was noted in a control F1 male. This finding was not made in 2,4-D-exposed adult F1 offspring with longer 2,4-D exposures, which included higher mg/kg/day exposures during critical windows of development, and there was no overall exposure-related impact on testes weight. Decreases in testis size are a spontaneous occurrence in adult Crl:CD(SD) rats (Pettersen *et al.*, 1996). There were no exposure-related changes in sperm parameters or histopathological changes in male reproductive organs. Overall, there was no indication of reproductive toxicity caused by 2,4-D in this study.

Potential endocrine toxicity

Findings are summarized in publication **Table 8**.

Estrogenicity/Anti-estrogenicity: There were no exposure-related effects on developmental landmarks, including AGD or age and body weight at vaginal opening. There were no effects on female reproductive organ histopathology in either P or F1 offspring, or on quantitative ovarian follicle counts in F1 offspring evaluated for this parameter. There were no significant, exposure-related changes in reproductive organ weights in P or F1 PND 22 females or in F1 adult males and females. As noted above, reproductive indices and litter size and pup survival were not affected by 2,4-D.

When uterine weights were examined at termination in P and F1 females, there was a suggestion of increased uterine weight in the high-dose group. The increases in uterine weights were not statistically significant compared to control, and showed high variability because the stage of the estrous cycle was not uniform at necropsy. There was general correlation with a higher (but not abnormally higher) incidence of 600 mg/kg/day females compared to control in proestrus or estrus at the time of necropsy. Proestrus and estrus are normal stages of the estrous cycle; these show the greatest fluid imbibition in the uterus, and therefore have higher uterine weights and hypertrophy compared to uteri at other stages of the estrous cycle. There were no signs of enlarged uteri noted on gross examination, and no abnormal findings on histopathological evaluation. Examination of HCD showed control uterine weights fell below HCD. Further, there was no suggestion of a compound-related effect on estrous cyclicity (animals were cycling normally, with no signs of persistent estrus, difference from control in cycle length, or of irregular cycles). There was no effect on time to mating (often prolonged if cycles are irregular). Based on these factors, the perceived increase in uterine weight is considered to reflect normal variability and not an exposure-related finding.

Androgenicity/Anti-androgenicity: In P adult males, decreased seminal vesicle and prostate weights were seen at $\geq$ 300 ppm; prostate weights were not statistically different from control. These findings were not considered exposure-related because both absolute and relative seminal vesicle and prostate weights in the control group exceeded the laboratory HCD ranges. Organ weights for the 800 and 300 ppm-exposed males were within the laboratory HCD range. Additionally, there were no associated histopathological findings. Differences in prostate and seminal vesicle weight seen in parental males were not reproduced in F1 offspring with longer 2,4-D exposures, including exposures during critical life stages. There were no exposure-related alterations in reproductive or accessory sex gland weights or histopathology in adult F1 males. Neither P nor F1 males (PND 139) showed effects on sperm parameters.

Decreased testis weights in 600 ppm PND 22 F1 weanlings, which lacked corresponding histopathological findings, were attributed to decreased body weights. Body weight findings correlated strongly with the testis weight findings, and resulted from artifactual differences in PND 22 male pup body weights introduced during group assignment, possibly exacerbated by 2,4-D-related toxicity and/or palatability issues in the 600 ppm group, which exceeds the TSRC. Note a previous feed restriction study indicated that weanling organ weights, including testes, decrease with alterations in body weight

(Carney *et al.*, 2004). This differs from results in adult male rats, where testes weights are conserved in the presence of body weight decrements (Chapin *et al.*, 1993). No effects on testis weights were seen in adult males.

There was a slight delay (+1.6 days) in F1 preputial separation at 800 ppm, which was attributed to high-dose body weight decrements and decreased growth during lactation and post-weaning. Body weight at the time of puberty onset was similar in 800 ppm males and controls. High-dose males weighed the same as controls 1.6 days later when puberty onset occurred, indicating that 800 ppm 2,4-D affected the rate of growth in peri-pubescent male rats. The magnitude of this delay (1.6 days with 7-8% differences in body weights on PND 28-42) was consistent with a 1.8-day delay in age at preputial separation in a feed restriction study with a 10% body weight decrement (Marty *et al.*, 2003).

Other androgen-sensitive endpoints including anogenital distance (AGD) and nipple retention in male F1 offspring were not altered. These endpoints are considered highly sensitive to anti-androgenicity (Clark, 1999; McIntyre *et al.*, 2001, Wolf *et al.*, 2002, Hotchkiss *et al.*, 2004). No effects were seen on other androgen-sensitive endpoints examined in the F1 generation.

Thyroid Assessment: Thyroid hormones (T3, T4, and TSH), thyroid weights and/or histopathology were evaluated at multiple life stages. There was no consistent pattern of effects on thyroid parameters across life stages. The only thyroid findings considered likely to be exposure-related were in high-dose satellite GD 17 dams given 600 ppm 2,4-D. These females had non-statistically significant decreases in T4 and T3 with a corresponding increase in TSH. Thyroid histopathological alterations were seen in 3 of 12 dams in this group. The histopathological findings were comprised of smaller thyroid follicles with small vacuoles in the colloid that were suggestive of colloid resorption. There were no adverse pathological alterations and thyroid changes were not observed in LD 21 main study dams, indicating that this effect was transient. The thyroid findings are considered adaptive, as decreased colloid is a normal feedback initiated response to slight fluctuations in circulating thyroid hormone levels. The NOAEL for this study for thyroid effects is the highest dose tested.

High-dose alterations in thyroid function during pregnancy are plausible given that high-doses of 2,4-D have been shown to compete with T4 for serum protein binding (Florsheim *et al.*, 1963 and Van den Berg *et al.*, 1991)) and because pregnancy affects all aspects of thyroid hormone economy (Larsen and Ingbar, 1992; Fukuda *et al.*, 1980). This adaptive change in thyroid function in GD 17 dams occurred only at 600 ppm, an exposure level above TSRC (nonlinear TK was particularly marked in GD 17 dams compared to non-pregnant adult females). There were no thyroid effects at lower dose levels in GD 17 dams and no statistically or biologically significant effects on thyroid endpoints at any of the other life stages examined.

Importantly, no developmental neurotoxicity (DNT) effects were observed in F1 males and females. Adverse effects on DNT parameters have been reported for compounds that are thyroid-active. The thyroid-active agent, 6-propyl-2-thiouracil (PTU), has been associated with decreased brain size, altered brain development (i.e., impaired neuronal migration and white matter hypoplasia) (Behnam-Rassoli *et*

al., 1991; Schoonover *et al.*, 2004; Lavado-Autric *et al.*, 2003; Shibutani *et al.*, 2009); effects not seen in the EOGRT study. There were also no changes seen in myelin deposition, which was reported as a concern by Duffard's group in Argentina (Duffard *et al.* 1996). Special staining was performed in the EOGRT study to evaluate potential effects on myelin deposition in developmentally and peri-natally exposed animals; none were observed.

Thyroid perturbations during development could affect motor activity, which would present as incoordination during clinical observations in pups and persistent hyperactivity (increased motor counts) in post-weaning animals during motor activity testing (Goldey *et al.*, 1995); these effects were not seen in the EOGRT study. Goldey *et al.* (1995) also reported that thyroid perturbations in offspring increased startle response amplitude in adults, again suggesting a hyper-reactive response. There were no effects on startle amplitude in the EOGRT.

Pituitary and Adrenal: Absolute and relative pituitary gland weights were significantly decreased in the 800 ppm F1 Set 3 males; however, this finding was not considered exposure related because of the extremely slight nature of the difference. Pituitary weights in P1 or F1 Set1a males and females or F1 Set 3 females were not statistically different. There were no exposure-related histopathological changes in pituitaries at any dose level or life stage.

There was no indication of alterations in adrenal function as there were no effects on absolute or relative adrenal gland weights and no exposure-related histopathological findings in adrenals at any 2,4-D exposure level. Adrenal gland weights and histopathology often are altered in the presence of steroidogenesis inhibitors (USEPA, 2007); therefore, these data do not support effects of 2,4-D on steroidogenesis.

In conclusion, there was no evidence of adversely altered endocrine function in a comprehensive EOGRT study of 2,4-D. Slight adaptive effects were seen on thyroid hormone homeostasis at the high dose in a single life-stage, at a dose exceeding the TSRC and not relevant for human risk assessment.

Rodwell and Brown, 1985

This study was a two-generation OPP 83-4 Guideline reproductive toxicity study in Fischer 344 rats. 2,4-D (97.5% purity) was administered in the diet at nominal dose levels of 0, 5, 20, and 80 mg/kg/day (30/sex/dose) for one full generation and at 0, 5, and 20 mg/kg/day for the second generation. The 80 mg/kg/day group was dropped after the first generation because it exceeded a maximum tolerated dose, based on excessive mortality among the F1b pups following a mis-dosing during gestation and lactation. The mis-dosing resulted in all groups of F1b dams and pups being exposed to greater than nominal doses; high-dose dam exposure was ≥ 100 mg/kg/day. Because of the mis-dosing and several deficiencies this study is scored a Klimisch score of 2.

The study followed standard FIFRA Guideline procedures of the time with the following exceptions:

- Two matings were conducted per generation

- Mating was conducted so one male with one female for a 10-day period; if there was no evidence of mating the female was re-paired with a proven male for an additional five days. The second mating of each generation tested different pairings.
- Pups dying before Day 4 were examined for anomalies externally and internally
- All F1b surviving pups were sacrificed at weaning; these pups, plus a random 10/sex from other dose levels were necropsied and kidney, liver, ovary, uterus, testes, epididymides, accessory sex glands and gross lesions evaluated histopathologically.

There were no treatment-related signs of toxicity or effects on survival of parental animals. At 80 mg/kg/day, decreased weight gain was seen during gestation. No decreased feed consumption was seen, except in dams producing the F1b pups. Fertility indices and mean pre-coital intervals were comparable for control and treated groups in three of the four matings in the study. The fertility index was slightly lower (not statistically significant) compared to control at 80 mg/kg/day in the mating to produce the F1b litters; the interval where the mis-dosing occurred.

The F1a litters showed a slightly increased number of stillbirths and decreased pup body weight; the F1b litters had a significantly increased incidence of stillbirths and marked decreases in pup survival (mortality in 68.3% of all pups born). The severity of the effect on the F1b pups was attributed to excessive maternal toxicity due to the unintentional overdose during gestation and lactation, although pup survival may also have been adversely impacted due to direct toxicity to the pups. There was no clear indication that lactation was adversely affected or any other indication that the pup deaths resulted from endocrine-related toxicity. At ≥ 20 mg/kg/day, pup body weights were slightly decreased (in F1b pups only). No adverse effects on pups were seen at 5 mg/kg/day.

Malformations observed in F1b pups at ≥80 mg/kg/day (as noted above likely more than 100 mg/kg/day dose to dams and clearly exceeding the TSRC) included generalized edema, mal-aligned sternebrae, and bent limb bones; variations included 14th ribs, bent ribs and decreased ossification of the vertebral arches (a limited number of F1b control pups were evaluated, therefore the exposure-relationship of these findings cannot be determined). The pattern of these developmental effects, however, is similar to that seen in some rat developmental toxicity studies with 2,4-D acid, esters, and salts at maternally toxic doses (Charles *et al.*, 2001) that also exceeded the TSRC and is not characteristic of the types of anomalies associated with endocrine disruption (e.g., urogenital malformations).

The high-dose effects on pup mortality, survival, and malformations are attributed to direct toxicity of 2,4-D on dams and/or pups; there is no evidence of an endocrine-related pattern. Additionally, no effects were noted on pre-coital length, fertility, or on histopathological evaluation of the accessory sex glands, epididymides, testes, ovaries, or uterus.

The gross and microscopic evaluation of uteri in the PND 28 offspring showed no evidence of imbibed uteri or uterine lining proliferation, even at the excessively toxic dose to the F1b weanlings. F344 pups tend to develop more slowly than do CD rats, and PND 28 pups although close to puberty are unlikely to be cycling, thus reducing uterine weight variability and making any estrogenic effect on the uterus more

readily observable. This evaluation therefore provides additional confidence that potential estrogenic effects do not result from 2,4-D exposure.

One finding suggesting possible endocrine toxicity was that the length of gestation was statistically significantly prolonged (by 1 day) in the production of the F1b pups at ≥80 mg/kg/day, compared with controls. Gestation may be prolonged because of difficulties in parturition, hormonal imbalance, or delays in implantation. The first alternative is unlikely because no evidence of dystocia (prolonged or difficult labor) was reported. Another possible explanation is that the females were at a more advanced age at the time of the second littering and may have been more susceptible to the test compound at the high dose which was well above the TSRC. The latter supposition is reinforced by the absence of this finding in the F1a littering (exposed at a lower dose) and is consistent with the absence of this finding in the Saghir *et al.* (2008a) range-finding study, in which a similar dose did not result in prolonged gestation. It should be noted that the F344 rat typical gestation period is 21-22 days; a day longer is likely of marginal biological significance. At most, the finding of prolonged gestation in the F1b litters provides equivocal evidence of a potentially treatment-related hormonal imbalance resulting from 2,4-D exposure at a dose significantly exceeding the RCST (and exceeding a classically defined MTD).

A second finding potentially showing endocrine toxicity was that the 80 mg/kg/day F1a pups showed a statistically significant change in sex ratio (109 males and 71 females), compared with the controls. This finding was not repeated in the F1b pups (at a higher dose) and is considered unlikely to be exposure related because of the lack of consistency. Additionally, there were no parallel exposure related effects on the sex ratio in the F1-extended one generation dietary toxicity study for 2,4-D (Marty *et al.*, 2010), or in the range-finding for the latter study (Saghir *et al.*, 2008a).

The Rodwell and Brown (1985) two-generation study was a complete and relatively well-designed assessment for its time, including adequate numbers of litters in each dose group for evaluation, with two litterings per generation. It was more comprehensive (because of the additional histopathological evaluations of weanlings) than required under the guidelines current at the time (OPP 83-4).

However, there were some study deficiencies to consider when evaluating these data for evidence of endocrine effects:

- Excessive toxicity occurred in the F1b animals at the high-dose level, likely due to an accidental mis-dosing of this group - actual dose exceeded 100 mg/kg/day; therefore only two dose levels were tested for the second generation.

- Additional parameters now recommended by guidelines on reproductive toxicity testing, which are relevant to assessment of potential endocrine disruption, were not included: estrous cyclicity; pituitary weights; developmental landmarks; and evaluations of sperm motility, morphology, or counts.

- F344 rats are not recommended for evaluation in the EDSP Tier 1 pubertal study guidelines (Sprague-Dawley rats are preferred) (US EPA, 2009h and US EPA 2009i).

- No thyroid-related parameters were evaluated as part of this study.

A summary of results for the Rodwell and Brown (1985) study is presented in publication **Table 9**. In the 2,4-D reproductive toxicity study, there were no treatment-related effects on pre-coital length, suggesting that the females had generally normal estrous cycling activity. There was no evidence of adverse treatment-related effects on the testes, which were evaluated microscopically in both adults and pups. The general reproductive success in this study suggests an absence of any potent androgen or estrogen-related toxicity; thyroid related endpoints were not assessed. Although pup weight gain and viability were compromised at ≥80 mg/kg/day, the severity of the effect on the F1b pups was attributed to the excessive maternal toxicity at this dose caused by inadvertent overdosing (to at least 100 mg/kg/day). There was no indication that lactation was adversely affected or other indication that the pup deaths resulted from endocrine-related toxicity. As discussed previously, the pattern of malformations seen at the high and maternally toxic dose tested generally paralleled findings in developmental toxicity studies of 2,4-D and related compounds at doses that far exceeded the TSRC, and were not typical for malformations relating to endocrine toxicity.

Developmental Toxicity

Rodwell, 1983

This study was a conventional OPP 83-3 Guideline developmental toxicity evaluation in F344 rats. The study design and results were published by Charles *et al.* (2001) along with the results of evaluations of selected 2,4-D esters and salts, which is summarized subsequently with relevant published data. 2,4-D (97.5% pure) was administered by gavage at doses 0, 8, 25, and 75 mg/kg/day from gestation day (GD) 6-15; rats were sacrificed and caesarian-sectioned on GD 20. No adverse effects were seen on corpora lutea, number of implantations, fetal survival, fetal sex ratio, urogenital malformations, any variations or malformation or fetal body weight after doses to the dams up to 75 mg/kg/day on gestational days (GD) 6-15. The high dose produced relatively slight maternal toxicity, although data from other studies would predict that the high dose by gavage would far exceed the renal clearance saturation threshold. There was no evidence of endocrine disruption (based on absence of uro-genital malformations in this study).

Hoberman, 1990

This study was a conventional OPP 83-3 Guideline developmental toxicity in New Zealand white rabbits. The study design and results were published by Charles *et al.* (2001) along with the results of evaluations of selected 2,4-D esters and salts, which are summarized subsequently with relevant published data. Rabbits were dosed by gavage with 0, 10, 30 or 90 mg/kg/day 2,4-D (97.5% pure) from GD 6-18. No adverse effects were seen on corpora lutea, number of implantations, fetal survival, urogenital malformations, any variations or malformation or fetal body weight after a dose to the dams of 90 mg/kg/day (HDT) on gestational days (GD) 6-18. The high dose produced maternal toxicity, including two abortions. The fetal sex ratio was altered at the high dose; this finding is considered unlikely to reflect endocrine toxicity because genotypic sex determination occurs prior to the start of dosing. Dantzler and Wright (2003) reported OAT-1 transporter in rabbit renal proximal tubules that actively transported p-aminohippurate, the classical renal organic anion transporter substrate. Consistent with

the presence of OAT-1 transporter in kidney, intravenous dosing of pregnant rabbits has demonstrated that renal clearance of 2,4-D was saturated at doses of 10 and 40 mg/kg (Sandberg *et al.*, 1996), suggesting that high-dose oral effects limited to 90 mg/kg/day likely occurred under conditions of non-linear toxicokinetic behaviors in rabbits. There was no evidence of endocrine disruption (based on absence of urogenital malformations in this study).

Subchronic and Chronic Studies

Schulze, 1991a

Male and female F344 rats were fed 2,4-D (96.1% pure) in the diet at 0, 1, 15, 100, or 300 mg/kg/day for approximately 13 weeks (10/sex/dose) in an OPP 82-1 Guideline study. The study design and results were published by Charles *et al.* (2006a) along with the results of evaluations of a selected 2,4-D ester and salt, which are summarized subsequently with relevant published data. Endocrine-related parameters in the Schulze (1991a) study included thyroid hormones (T3 and T4), ovary, vagina, testes with epididymides, thyroid, adrenal and pituitary weights and histopathology, and uterus histopathology.

Females fed 300 mg/kg/day exhibited depressed activity. At 13 weeks, mean body weight and body weight gain of male rats fed 100 or 300 mg/kg/day were decreased (7.5% and 23%, respectively). Mean body weights of female rats fed 300 mg/kg/day at 13 weeks were decreased (28%). The two high doses in this study clearly exceeded the renal clearance saturation threshold, and the high dose exceeded a classically defined MTD. Systemic toxicity was marked at 300 mg/kg/day. Renal toxicity was present at 100 and 300 mg/kg/day based on renal weight changes and/or renal histopathology. High dose effects also included effects on eyes, hearts and lungs.

Findings potentially related to endocrine modulation were limited to high systemically toxic doses above the TSRC:

- T4 was depressed in the males and females at 100 or 300 mg/kg/day at weeks 6 and 13. T3 levels were decreased in females fed 100 and 300 mg/kg/day at 13 weeks. T3 was depressed in the males fed 300 mg/kg/day at 13 weeks. (TSH was not evaluated). Males had higher absolute thyroid/parathyroid weights at 300 mg/kg/day and relative weights at 100 and 300 mg/kg/day. Females at 300 mg/kg/day had higher relative thyroid/parathyroid weights. Increased follicular cell hypertrophy was observed in the thyroid glands of females at 300 mg/kg/day (8/10). No exposure-related findings were noted in the thyroid of males.

- Grossly, the cortex of the adrenal glands was pale in males fed 300 mg/kg/day (5/10). No changes were noted in the medulla. Males fed 100 and 300 mg/kg/day had dose related higher relative adrenal gland weights. Females fed 300 mg/kg/day had lower absolute adrenal gland weights. There was an increased incidence of hypertrophy in the zona glomerulosa of the adrenal gland cortex in males fed 300 mg/kg/day, and females fed 100 and 300 mg/kg/day.

- The ovary weight was increased at 300 mg/kg/day; there was no histopathological correlate.
- The testes at 300 mg/kg/day were soft (8/10) and small (7/10). Epididymides showed similar findings at this dose. Lower absolute and relative testes weights were noted at 300 mg/kg/day. Increased testicular atrophy was seen histopathologically at this dose.
- Females fed 300 mg/kg/day had lower absolute and relative pituitary weights. Males fed 300 mg/kg/day had lower relative pituitary weights. There was no histopathological correlate to this finding in either sex and it is considered unlikely to be exposure-related.
- Histopathological changes were not observed in the following tissues: pituitary, parathyroid glands, epididymides, ovary, uterus, and vagina.

Gorzinski *et al.*, 1981a

Male and female F344 rats were fed 2,4-D (97.3% pure) in the diet at doses approximating 0, 15, 60, 100, or 150 mg/kg/day for approximately 13 weeks (15/sex/dose) in an OPP 82-1 Guideline study. Endocrine relevant parameters included thyroid hormone (T4); ovary, testes with epididymides weights and histopathology; and thyroid, pituitary, adrenal, prostate, seminal vesicle, mammary gland, oviduct and uterus histopathology.

Males fed 150 mg/kg/day showed a 5% decrease in body weight. Females fed 100 and 150 mg/kg/day had a 5-7% and 10% decrease in body weight, respectively. The three high doses exceeded the T5RC for both males and females. Systemic toxicity was apparent at 150 mg/kg/day, including renal toxicity characterized by renal weight changes, slight swelling of the kidneys noted during gross pathology, and/or renal histopathology.

Findings potentially related to endocrine modulation were:

- Serum T4 was reduced in a dose dependent manner in females fed 100 and 150 mg/kg/day. No statistically significant reduction was noted in males. T3 and TSH were not measured. Thyroid weights were not measured. No histopathological changes in the thyroids were noted in either sex.
- The relative testes weight was decreased in male rats fed 100 and 150 mg/kg/day. No correlating histopathological changes were observed.

No histopathological changes were noted in the adrenal glands, ovary, uterus, pituitary glands, epididymides, seminal vesicles, coagulating gland, parathyroid glands, mammary glands or oviducts.

Gorzinski *et al.*, 1981b

Male and female F344 rats were fed 2,4-D (100% pure) in the diet at dietary concentrations equivalent to 0, 15, 60, 100 or 150 mg/kg/day (15/sex/dose) in a non-Guideline sub-chronic toxicity study with purified test material examining endocrine relevant endpoints of thyroid hormone (T4) and testes

weight. Note although this is scored a Klimisch 3 because of very limited evaluations it is summarized here because it provides supplemental information to Gorzinski *et al.* (1996a).

Mean body weight of females was decreased (7.6%) at 150 mg/kg/day and body weight of males was decreased (4.6 and 7.7%) at 100 and 150 mg/kg/day, respectively. The two highest doses exceed the threshold for renal saturation for both males and females. Renal toxicity was noted in males and females fed 100 and 150 mg/kg/day based on renal weight change and swelling noted grossly for kidneys. Renal histopathological changes were seen in males at 150 mg/kg/day.

Findings potentially related to endocrine modulation were limited to decreased T4 in females fed 60, 100, or 150 mg/kg/day. T4 levels in males were variable. There were no exposure-related effects on testes weights.

Note no organ weights or histopathology was performed on endocrine-sensitive organs such as the adrenal glands, thyroid glands, ovary, uterus, pituitary glands, epididymides, seminal vesicles coagulating glands, parathyroid glands, mammary glands or oviducts. Thus, this study is limited in the amount of information relevant to assessing endocrine modulation.

Jeffries *et al.*, 1995

In this rat chronic toxicity/oncogenicity study, male and female F344 rats were administered 2,4-D (96.45% pure) in the diet at 0, 5, 75, or 150 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). Endocrine-relevant endpoints included: thyroid hormone measurement (T4); ovary, testes, thyroid, adrenal weights and histopathology; and pituitary, oviducts, cervix, uterus, vagina, mammary gland, epididymides, seminal vesicle, coagulating gland and prostate histopathology. The only deviation from the guideline at the time the study was performed is that the high dose female dose exceeded an MTD.

Mid-and high doses exceeded the TSRC and the high dose females exceeded a classically defined MTD based on bodyweight gain depression. There was no exposure-related mortality. There was no indication of increased tumor incidence, endocrine or otherwise. Statistically decreased body weight was reported for males at 150 mg/kg/day and females at 75 mg/kg/day (both at 12 months and 24 months). Primary effects were seen in the eyes, liver, and lungs of male and females at 150 mg/kg/day, the heart of males at 150 mg/kg/day, and the lungs of females at 75 mg/kg/day. Secondary effects were considered related to the decreased body weight and feed consumption at 75 and 150 mg/kg/day. At 12 months, a non-statistically significant increased incidence of very slight or slight degeneration of the descending proximal tubule of the kidney was indicated at 75 and 150 mg/kg/day. A high incidence of increased panlobular hepatocytes size often accompanied by altered tinctorial properties was reported for both males and females at 150 mg/kg/day, an increased incidence of very slight and decreased slight bile duct hyperplasia with or without inflammation was reported for males at 150 mg/kg/day. Analyses of severity showed no statistically significant changes.

The following findings represent potentially endocrine related effects. It should be noted these were seen only above the TSRC and in the presence of significant other systemic toxicity.

- At 12 and 24 months, thyroid hormone levels (T4) were statistically decreased at 75 and 150 mg/kg/day in both males and females.
- Statistically increased thyroid weights (absolute and relative) were seen at 12 and 24 months in males at 150 mg/kg/day and females at 75 and 150 mg/kg/day; this increase was attributed to thyroid masses. No clear histopathological correlate was evident for these masses, however, females at 150 mg/kg/day at 24 months had an increased incidence of parafollicular cell nodular hyperplasia compared to controls. Parafollicular cell hyperplasia is not typically observed in response to decreased circulating thyroid hormones, and the exposure-relationship of this finding is questionable. [Note that tissue accountability at 24 months was poor for female thyroids due to a labeling problem during processing. The individual animal identification for 35 female thyroid/parathyroid glands was lost during the preparation of these tissues for histological evaluation during the dehydration process as a result of removal of the identifying ink from the tissue containers. This resulted in 8, 9, 9, and 10 thyroid glands missing from the 0, 5, 75, and 150 mg/kg/day groups respectively. However, because the misidentified tissues were spread across dose groups and adequate numbers of appropriately labeled tissues were available for evaluation, this deviation had no effect on interpretation of the study outcomes.]
- At 12 months, females showed a very slight decreased secretory material in the epithelial cells of the thyroid at 150 mg/kg/day compared to control. This change suggests an adaptive response to decreased circulating T4 and is considered likely to be exposure-related. At 24 months in all modes of death (early termination through 24 months), an increased incidence of focal cystic dilatation was observed in follicles of the thyroid gland of females at 75 and 150 mg/kg/day.
- At 24 months, statistically decreased weight of ovaries (absolute and/or relative) at was found at 75 and 150 mg/kg/day; there was no histopathological correlate and this finding is considered likely not to be biologically significant but related to decreased body weight.
- At 12 and 24 months, statistically decreased weight of testes (absolute) was seen at 150 mg/kg/day. Relative testes weights were also decreased at 24 months. No exposure-related histopathological changes were seen in the testes.
- At 24 months in all modes of death (early termination through 24 months), a decrease in very slight focal or multifocal area of altered cells was found in the cortex of the adrenal glands of females at 150 mg/kg/day. No other histopathology was indicated; decreased incidence of primary benign adenoma in the cortex of the

adrenal glands was reported in females at 150 mg/kg/day. This change may be exposure related, however, cannot be characterized as adverse. There were no gross findings in the adrenals at either sacrifice interval, and no exposure-related histopathological changes in the adrenals at the 12-month sacrifice.

- At 24 months in all modes of death (early termination through 24 months), there was a decreased incidence of pituitary mass/nodules in both males and females (not statistically significant. At 24 months (all modes of death), there was a decreased incidence in benign adenomas of the *pars distalis* in the pituitary, which was relatively slight in males, but marked in females at 150 mg/kg/day. The incidence in males was (19/50 (38%), 15/32 (47%), 12/28 (43%), 9/49 (18%)) and females was (21/50 (42%), 15/39 (39%), 9/21 (43%), 1/50 (2%)) at 0, 5, 75, and 150 mg/kg/day, respectively. The decrease in the incidence of this relatively common likely estrogen responsive tumor in females is considered related to marked weight loss at the high dose exposure.
- Gross pathology at 24 months showed a decreased incidence of mammary left inguinal hyperplasia in female rats (16/50, 24/50, 12/50, and 1/50) at 0, 5, 75, and 150 mg/kg/day, respectively. This was not noted in males. At 24 months in all modes of death (early termination through 24 months), a decreased incidence of mammary gland hyperplasia often accompanied by duct ectasia was reported at 150 mg/kg/day (26, 14, 11, 9 in males and 36, 33, 16, 1 in females) at 0, 5, 75, and 150 mg/kg/day, respectively. This decrease may be attributable to decreased estrogens but is most likely related to the marked weight loss at the high dose exposure. No gross or microscopic findings were made in the mammary glands at the 12 month sacrifice.

Other endocrine related tissues showed no exposure related gross or microscopic findings: parathyroid glands, preputial or clitoral glands, prostate, seminal vesicles, oviducts, uterus and vagina.

<u>Schulze, 1991b</u>

Male and female B6C3F1 mice were administered 2,4-D (96.1% pure) in the diet at 0, 1, 15, 100, or 300 mg/kg/day for approximately 13 weeks (10/sex/dose). Endpoints evaluated relevant to endocrine effects include thyroid hormone (T4), ovary, thyroid, pituitary, adrenal and testes weights and histopathology, and uterus and epididymides histopathology.

There were no significant differences between treated animals from control animals in overall body weight or body weight gains. At 13 weeks, the livers of males and females at 300 mg/kg/day had elevated nuclear hyperchromatism in periportal hepatocytes (8/10 animals/sex). The absolute kidney weights decreased (non-statistically) at 300 mg/kg/day for males. Tubular degeneration was observed in 9/10 male mice at 300 mg/kg/day; this was not observed in

females. Other renal changes in males at 300 mg/kg/day include: karyomegaly, loss of brush border, and decreased size of tubular lining. The absolute and/or relative kidney weight in females was increased at 100 or 300 mg/kg/day. The lack of histopathological correlation in the females makes the biological significance of the slight changes in the kidney weight questionable.

Findings potentially related to endocrine modulation include:

- Thyroxine (T4) in males and females was non-statistically significantly decreased at 100 and 300 mg/kg/day. There was a dose-dependent decrease in the males. There were no changes in thyroid weight or histopathology and the T4 decrease, although likely exposure-related, is not considered an adverse effect.
- The absolute weights of the adrenal glands were slightly but statistically increased at 1, 15, and 100 mg/kg/day (with an apparent increase at 300 mg/kg/day) for females; no changes were reported for males. The increase was similar and very slight at all doses except 300 mg/kg/day. The relative weights of the adrenal glands were statistically elevated compared to controls for 1, 15, and 300 mg/kg/day (with an apparent increase at 100 mg/kg/day). No dose dependence was observed. There were no histopathological findings in the adrenal for either sex at any dose, and the non-dose related adrenal weight changes are considered unlikely to be biologically significant or exposure-related.

There were no significant differences in incidences of gross pathology of treated animals from control animals were found in the following tissues: pituitary, adrenal glands (cortex and medulla), thyroid glands, parathyroid glands, liver, kidney, testis, epididymides, ovary, uterus, or mammary glands. No significant differences in absolute or relative organ weights of treated males or females were identified in the following tissues: ovary, liver, testes, thyroid glands or pituitary glands. No histopathological changes in treated animals were reported for the following tissues: pituitary, adrenal glands (medulla and cortex), thyroid glands, parathyroid glands, testes, epididymides, ovary, or uterus. This study does not demonstrate any adverse endocrine-related effects in mice at doses below 100 mg/kg/day in a 90-day study. Although direct data are lacking, it is likely that high-doses of 2,4-D are above the TSRC of 2,4-D in mice. Both rats and mice express comparable levels of OAT-1 transporter mRNA in the kidney (Buist and Klaassen, 2004).

Stott *et al.* 1995a

In this mouse oncogenicity study, female B6C3F1 mice were administered 2,4-D (96.4% pure) in the diet at 0, 5, 150, or 300 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). There were no long term changes in body weight or body weight gains of mice exposed to 2,4-D. At 12 months absolute and relative kidney weights increased at 150 and 300 mg/kg/day (13 and 16%). At 24 months the relative kidney weight increased at 150 and

300 mg/kg/day (11 and 20%). At 24 months, histopathological effects on kidneys included hypercellularity of the descending portion of the proximal tubule and degeneration/regeneration of the cortical tubules at 150 and 300 mg/kg/day and mineralization of renal tubules at 300 mg/kg/day.

There were no effects on endocrine responsive tissues (based on histopathological evaluation) in this study at 12 or 24 months, including: adrenals, cervix, liver, mammary glands, ovaries, oviducts, parathyroids, pituitary, thyroids, uterus, and vagina.

Stott *et al.* 1995b

In this mouse oncogenicity study, male B6C3F1 mice were administered 2,4-D (96.4% pure) in the diet at 0, 62.5 or 125 mg/kg/day for approximately 12 months (10/sex/dose) or 24 months (50/sex/dose). There were no changes in body weight or body weight gains reported. There was no exposure related incidence of neoplasms. At 24 months, kidney weight increased at 62.5 and 125 mg/kg/day and histopathological lesions of the kidney were apparent at both 12- and 24-months. In the liver at 24 months, there was an increase in multifocal areas of aggregates of reticuloendothelial cells frequently adjacent to degenerative or necrotic hepatocytes at 125 mg/kg/day.

There were no exposure-related endocrine related effects in this study. At 12 months absolute and relative testes weights were not different from control. At 24 months, testes relative organ weight was increased at 125 mg/kg/day (6%). However, there were no histopathological correlate to the weight change, and the change was slight in nature, and absolute testes weights were not statistically significantly increased. Therefore this finding is not considered exposure-related.

No changes in gross pathology were reported at 12 or 24 months for the following tissues: kidneys, liver, seminal vesicles (12 months only), testes, adrenals (24 month only), and epididymides (24 months only). No histopathological changes occurred in the following tissues at 12 or 24 months: adrenals, coagulating glands, epididymides, parathyroid, prostate, seminal vesicles, testes, or thyroid gland.

Schulze, 1990

In this sub-chronic study, 2, 4-D (96.1% pure) was administered to dogs via capsule at 0, 0.3, 1, 3 or 10 mg/kg/day for 13 weeks. Endocrine related endpoints evaluated included: thyroid hormones (T3 and T4), ovary, thyroid and testes (with epididymides) weights and histopathology, and adrenal, pituitary and uterus histopathology.

There was severe systemic toxicity at the highest dose tested (10 mg/kg/day) (HDT): including mortality, weight loss, renal pathology and altered BUN and creatinine. At 3 mg/kg/day renal pathology and altered BUN and creatinine were evident. The 10 mg/kg/day dose clearly exceeded a Maximum Tolerated Dose in dogs.

Interestingly, despite the severe toxicity in this study there were no exposure related effects on T3 or T4 levels, and no exposure related effects on thyroid hormones (T3 and T4), weights or histopathology. This stands in contrast to the rodent data, and highlights the susceptibility of the rat to changes in thyroid homeostasis. In addition, although thyroid toxicity was absent dogs, 2,4-D serum concentrations likely were substantially higher in dogs than in rats. Administration of a single oral dose of 5 mg/kg 2,4-D to either dogs or rats resulted in a serum AUC that was 232-fold higher in dogs compared to rats (vanRavenzwaay *et al.*, 2003).

Decreased testes weight and testicular findings were present at the high dose. This finding will be discussed in conjunction with the other dog sub-chronic study below.

Dalgard, 1993a

A second subchronic dog study was conducted at the same laboratory as the Schultze (1990) dog study using 2,4-D (96.7% pure) administered in the diet. The study design and results were published by Charles *et al.* (2001b) along with the results of evaluations of selected 2,4-D esters and salts, which are summarized subsequently with relevant published data. In the Dalgard (1993a) study, concentrations were administered to achieve targeted dose levels of 0, 0.5 1, 3.75 or 10 mg/kg bw/day; after 8 weeks the 10 mg/kg/day dose was lowered to 7.5 mg/kg bw/day based on excessive toxicity. Endocrine relevant parameters evaluated in this study included adrenal, ovary, pituitary, thyroid and testes (without epididymides) weights and histopathology, and epididymides, prostate, mammary gland, uterus and vagina histopathology.

Body weight gain in this study was decreased at the mid and high dose but not statistically significantly. At 7.5 mg/kg/day, testes weight was decreased and relative but not absolute thyroid weight was increased. No effects were seen on any other endocrine-relevant parameter. The toxicological significance of the thyroid findings is not clear given the absence of any correlating histopathological findings, the lack of absolute weight increase in the thyroid, and the lack of effects on thyroid hormone levels in the prior sub-chronic dog study at a higher dose level. Based on these factors, the increased relative thyroid weight is considered unlikely to be exposure related.

Regarding the testis weight decreases seen in both subchronic studies, and the dose-related incidence of testicular lesions in the Schulze (1990) study, the evidence suggests that a high number of juvenile dogs in both subchronic studies may have contributed to the observed effects. (The age of the dogs is not defined in the study reports; however a subsequent paper by Charles *et al.* (1991b) indicates the dogs in the Dalgard (1993a) study were 4-6 months of age at study initiation.) A comparison of body weights and testes and thyroid histopathology in both sub-chronic dog studies (see **Table S8**), leads to the conclusion that the majority of the dogs in both the sub-chronic studies were juvenile animals on the low end of the stated age range. (Both studies were conducted at the same laboratory during a similar time frame.)

Comparison of the body weights of the dogs at study initiation supports that the dogs were of a similar age and that they were young (mature male beagle adult weight is 10-12 kg). Testicular giant cells and hypospermatogenesis are common background lesions in beagles (up to 30% per Rehm, 2000). Immature dogs (less than 9 months of age) have been reported to have control incidences as high as 75% of both decreased testes weight and hypospermia (Goedken *et al.*, 2008). In the Goedken *et al.* analyses of data from a large population of control dogs, atrophic/hypoplastic tubules were seen in 26.3% of all dogs with 25%–40% of dogs under twelve months old having this finding, decreasing with increased age to 14%–17% in dogs twelve to thirty-six months old. Additionally, the high incidence of juvenile prostate findings in the second study (prostate was not evaluated in the first study) supports that the dogs were immature. It seems likely that the decreased testes weights in both sub-chronic studies and histopathological findings in the testes of the dogs in the 1990 study are an artifact related to the young age of these animals, although it is possible that the high (and lethal) dose of 10 mg/kg/day contributed to delayed development. Supporting the possibility of artifact, a chronic study in dogs (Dalgard, 1993b, discussed below) showed no exposure-related effects on testes weights or testicular histopathology following a one year exposure to 2,4-D at a high dose level (10 reduced to 7.5 mg/kg/day) generally equivalent to the high dose in the prior sub-chronic studies.

Dalgard, 1993b

In a chronic (1 year) dog study, 2,4-D (96.7% pure) was administered in the diet at target doses of 0, 1, 5 and 10 mg/kg bw/day. After 8 weeks the high dose was dropped to 7.5 mg/kg/day from 10 mg/kg/day. The study design and results were published by Charles *et al.* (1996b) along with the results of evaluations of selected 2,4-D esters and salts sub-chronic toxicity studies. Body weight and feed consumption were measured; compound intake calculated and dietary concentrations confirmed. Potential endocrine related parameters in this study included organ weights and histopathological evaluation of adrenals, testes, ovaries, pituitary and thyroid, and additional histopathological evaluation of epididymides, prostate, uterus, vagina, and mammary gland.

Body weight was markedly decreased in the high dose group, severely in females before lowering the high dose. Renal pathology and altered BUN and creatinine were evident at 10/7.5 and 5 mg/kg/day. The NOAEL was 1 mg/kg/day. There were no effects on endocrine parameters. Testicular and ovarian weights and histopathology were not affected, nor were thyroid weights or histopathology.

The absence of effects on testes weights and histopathology supports that the sub-chronic study findings were primarily related to the immature age of the test animal, as the high dose in the chronic study was similar to that in the sub-chronic studies.

As noted in the publication, the dog is not regarded as a relevant species for human risk assessment given the substantial differences in 2,4-D toxicokinetics in this species relative to rats and humans (Timchalk, 2006; vanRavenzwaay *et al.*, 2003). Thus, 2,4-D toxicity in dog is only relevant to assessment of potential environmental mammalian species impacts. It is also potentially relevant in that it supports the unique susceptibility of rodents (especially rats) to perturbation of thyroid hormone balance.

B. Studies Identified in the Published Literature

Developmental Studies

Charles *et al.*, 2001

Developmental toxicity studies in rats and rabbits, performed with the following chemicals: 2,4-D acid, 2,4-D dimethylamine salt (2,4-D DMA), and 2,4-D 2-ethylhexyl ester (2,4-D EHE); 2,4-D diethanolamine salt (2,4-D DEA); 2,4-D isopropylamine (2,4-D IPA); 2,4-D triisopropanolamine (2,4-D TIPA); 2-butoxyethyl ester (2,4-D BEE); and 2,4-D isopropyl ester (2,4-D IPE) were summarized in a publication by Charles *et al.* (2001). These were GLP studies designed to comply with USEPA Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA) testing guidelines. The study methods and results are reviewed in detail in Charles *et al.* 2001; the studies are briefly summarized below (with the exception of the 2,4-D acid rat (Rodwell, 1983) and rabbit (Hoberman, 1990) data, which are included with the regulatory studies of 2,4-D, based on the data in the study reports).

Rat

The rat developmental toxicity studies comply with OPP 83-3 guidelines and are considered to meet Klimisch criteria 2, because of a lack of sufficient detail in the published studies, although it should be noted current guidelines require a longer exposure period. In the rat developmental toxicity studies, groups of 25 to 35 bred female rats/group were gavaged daily on GD 6-15. Sprague-Dawley rats were used in these studies, except for the study on 2,4-D acid, which used Fischer 344 rats. (The 2,4-D acid developmental study (Rodwell, 1983) is reviewed in the discussion of regulatory toxicity studies; this study is scored a Klimisch 1.) The dose levels for each compound are indicated in **Table 12** in the publication, but range from 8-150 mg/kg/day expressed as 2,4-D acid equivalents (ae). Separate control groups for each compound were administered the appropriate vehicle. The 2,4-D ae values will be used in the following discussion.

Dosing was during the period of major organogenesis. Dams were observed twice daily, and maternal body weights and feed consumption were recorded. Surviving dams were euthanized on GD 20 and examined grossly, and kidney and liver weights recorded. Dams dying or sacrificed moribund before GD 20 had gross necropsies performed. At cesarean section, gravid uterine weight, number of corpora lutea, number and position of implantations, resorptions, and live or dead fetuses were recorded. Uteri with no visible implantations were stained for evidence of early resorptions. Each fetus was individually identified, weighed, sexed, and given a gross examination for external malformations/variations. Approximately one-half of the rat fetuses in each litter were evaluated for visceral malformations/variations, with heads examined for craniofacial defects. The remaining rat fetuses in each litter were processed for evaluation of skeletal alterations.

Developmental toxicity with the various 2,4-D salts and esters was observed only at dose levels causing maternal toxicity and was dose related. In general, maternal body-weight effects in rats began to be apparent at dose levels of 30 mg/kg/day. This dose by gavage has been shown to exceed the renal

clearance saturation threshold. At 30 mg/kg/day, body-weight gain was significantly depressed with 2,4-D EHE, 2,4-D EHE and 2,4-D IPE. At dose levels ≥90 mg/kg/day ae, clinical signs of toxicity were reported with 2,4-D DEA, 2,4-D DMA, 2,4-D IPA, and 2,4-D BEE, and maternal mortality was seen with 2,4-D TIPA. The no-observed-adverse-effect level (NOAEL) in dams across the studies ranged from 8 to 17 mg/kg/day.

There were no treatment-related effects on litter size, resorption rates, or fetal sex ratios. Developmental effects generally occurred only at the highest dose levels tested (HDT). Significantly decreased fetal-body weights were seen with 2,4-D DEA, 2,4-D DMA, 2,4-D TIPA, 2,4-D EHE, and 2,4-D IPE at >90 mg/kg/day (the HDT of each compound). Statistically significant treatment-related increases in fetal variations observed with 2,4-D DEA, 2,4-D DMA, 2,4-D TIPA, 2,4-D BEE, and 2,4-D IPE at doses >90 mg/kg/day included slightly delayed skeletal ossification and the presence of extra ribs, either cervical or lumbar. With 2,4-D EHE, a statistically significant increase in the incidence of incomplete or unossified sternebrae was the only effect noted in fetuses at 30 mg/kg/day and above. A statistically significant increase in the incidence of external and visceral malformations and skeletal anomalies was seen only at the high-dose level of 2,4-D TIPA (175 mg/kg/day), consisting of malformations of the eyes and rib variations (wavy, fused). This dose level exceeded an MTD, with maternal toxicity including mortality and severely decreased maternal weight gain. The types of developmental findings were not those typically characteristic of endocrine system disruption.

Rabbit

Groups of 18 to 24 inseminated adult female New Zealand white rabbits were administered the test materials by oral gavage once daily on either GD 6-18 or 7-19. Dose levels are shown in the preceding table but range from 10 to 75 mg/kg/day expressed as 2,4-D ae. Separate control groups were administered the appropriate vehicle.

The studies complied with the OPP 83-3 Test Guidelines. The developmental studies in rabbits (except for the study on 2,4-D acid (Hoberman, 1990) which is scored a 1) are considered Klimisch criteria 2, primarily because of a lack of some detail in the presentation, although it should be noted current guidelines require a longer exposure period. Additionally, the study on 2,4-D DMA was scored a 2 because maternal toxicity (not clearly dose related) limited the number of litters available for evaluation. However, developmental toxicity was considered adequately characterized.

Does were observed twice daily, and maternal body weights recorded at intervals during gestation. Surviving does were killed on GD 28 or 29; gross postmortem examinations were performed on all females (including deaths or moribund sacrifices). At cesarean section, gravid uterine weight, number of corpora lutea, number and position of implantations, resorptions, and live or dead fetuses were recorded. Uteri with no visible implantations were stained with an ammonium sulfide solution to detect early resorptions. Each fetus was individually identified, weighed, sexed, and given a gross examination for external malformations/variations. All fetuses were examined by dissection for evidence of visceral alterations and processed for evaluation of skeletal alterations.

In rabbits, dose-related maternal toxicity was seen with 2,4-D salts and esters at doses at or above 30 mg/kg/day ae. With higher doses (75-90 mg/kg/day), more severe maternal effects were noted. Clinical signs of toxicity accompanied by maternal body-weight losses, and in some cases, significant morbidity and/or mortality (2,4-D DMA, 2,4-D IPA, 2,4-D TIPA, and 2,4-D BEE) was observed.

In rabbits, embryonic and fetal development was essentially unaffected, even at maternally toxic doses. There were no effects on maternal reproductive measures such as litter size, number of resorptions, or on fetal body weights. Increased 7th cervical ribs were observed in the group exposed to 2,4-D DEA at 40.6 mg/kg/day, a dose level that also produced clinical signs of toxicity and decreased maternal body weight gain. Although statistically significant compared to control, it should be noted this is a finding with a high background incidence. There was no evidence of teratogenicity in rabbits with 2,4-D acid or its salts or esters. Additionally, there was no evidence of urogenital anomalies characteristic of endocrine disruption in these studies.

The types of findings in both the rat and rabbit developmental toxicity studies are characteristic of those associated with maternal toxicity, and do not suggest endocrine-mediated effects. Adverse effects on the developing rat and rabbit fetuses exposed *in utero* to 2,4-D were observed only at dose levels that produced maternal toxicity, with increasing dose levels that exceeded he TSRC (established for rats only, but predicted for rabbits based on similarities in their renal transport capability), causing increasingly more severe maternal effects, with concomitant effects on the developing fetus. Further, the studies do not provide any specific evidence of endocrine disruption. The majority of the 2,4-D forms were not associated with major malformations; the single exception is found in the rat study with 2,4-D TIPA at a very severely maternally toxic dose. This review of developmental findings for 2,4-D and its forms is included in the WoE because it shows the commonality of findings for this group of related compounds and does not predict endocrine disrupting activity for any of the compounds.

Dinamarca *et al.*, 2007

Eight-week old, ICR/Jcl mice were mated, and were subsequently administered 2,4- D as a "pure compound" (purity unspecified) or as a commercially available formulation available in Chile (unspecified) in drinking water at concentrations providing mg/kg/day doses of 0, 0.01, 0.10 or 100 mg/kg/day from GD 0-9. GD 0 was designated as the day a copulatory plug was observed. There were 10-13 mice/group. Dam body weights were recorded at GD 0, 6 and 9. Water consumption was measured, and mice were monitored for behavior. Feed consumption was stated to have been observed; it was not clear if this was measured. Mice were bled at GD 9 for biochemical evaluations and caesarian-sectioned. Ovaries were evaluated for numbers of corpora lutea and uterine horns were evaluated for number of implantation sites, resorptions and live embryos.

Oxidative stress was assessed by the determination of catalase activity, thiobarbituric reactive species (TBARs) and total antioxidant capacity (TAC). Carbon tetrachloride was administered subcutaneously at 3 ml/kg on GD8 as a positive control.

There were no signs of maternal toxicity nor differences in body weight gain between the dosed groups and the control. Numbers of corpora lutea, implantation sites, resorptions and live embryos were similar between the dose groups and control.

TAC values were significantly decreased only after administration of 100 mg/kg/dose 2,4-D of both "pure" 2,4-D and the formulated test material, suggesting that non-enzymatic antioxidant defenses may be depleted at that dose level which likely exceeds the TSRC in mice. There are some reservations regarding this conclusion because it is not clear how the TAC was assessed; multiple pathways are involved. Catalase activity and TBARS were not changed by exposure.

This study includes a fairly complete description of methods and appropriate analyses of results. One flaw is that details on the purity and source of the "pure" 2,4-D are not provided. The dose spacing in this study was designed to address the low dose hypothesis proposed by Cavieres *et al.* (2002). (The Cavieres *et al.* work is summarized in Supplementary Appendix VII.) The Dinamarca *et al.* study demonstrated that the finding of decreased implantations in mice exposed to 2,4-D in the Cavieres *et al.* (2002) report could not be replicated, even with an exposure period correctly designed to explore this possibility

Male Reproductive Toxicity

Lamb *et al.*, 1981a

In this study, male C57BL/6N mice were dosed via diet for 8 weeks with various concentrations of 2,4-D; 2,4,5-T; and TCDD. Approximate daily exposures were 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 2.4 micrograms per kilogram per day (µg/kg/day) TCDD; 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 0.16 µg/kg/day TCDD; and 20 mg/kg/day 2,4-D, 20 mg/kg/day 2,4,5-T, and 1.2 µg/kg/day TCDD for groups II, III, and IV respectively. Group I (control) received untreated diet. Fertility, sperm number, motility, and morphology were evaluated. Somatic cell sister chromatid exchange frequencies were also evaluated in mice injected intraperitoneally with single doses of similar chemical mixtures. No significant effects were observed in the male mice exposed to 2,4-D in the above mixtures compared to the control groups. Thus, this study provides no evidence of endocrine-disrupting activity of 2,4-D. Its primary weaknesses are that it tests only mixtures, however, this would not be expected to confound the negative results, and the dose and purity of the 2,4-D are defined. The high dose of 2,4-D would be anticipated to approximate the TSRC in mice so exposure to 2,4-D was not limited by the toxicity of the other mixture components.

Lamb *et al.*, 1981b

In this study, male C57BL/6N mice were dosed via diet for 8 weeks with various concentrations of2,4-D, 2,4,5-T, and TCDD. Approximate daily exposures were 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,5-T, and 2.4 µg/kg/day TCDD; 40 mg/kg/day 2,4-D, 40 mg/kg/day 2,4,S-T, and 0.16 µg/kg/day TCDD; and 20 mg/kg/day 2,4-D, 20 mg/kg/day 2,4,5-T, and 1.2 µg/kg/day TCDD for groups II, III, and IV respectively. Group I (control) received untreated diet. Males were mated with untreated female mice. Females

were either cesarean sectioned at gestation day 18 for evaluation of the fetuses, or allowed to deliver and rear their pups until weaning at postnatal day (PND) 21 for evaluation of offspring birth weight and viability. There was no residual impact on reproductive performance of the treated males in this study. Development and survival of fetuses and pups in the offspring of treated groups were similar to that of the control mice, and no effect of these mixtures was apparent. The study provided no evidence of male-mediated reproductive toxicity or of endocrine disrupting activity of 2,4-D for these mixtures of chemicals. Its primary weakness is that it tests only mixtures; however, this would not be expected to confound the negative results. The high dose of 2,4-D would be anticipated to approximate the TSRC in mice so exposure to 2,4-D was not limited by the toxicity of the other mixture components.

Subchronic and Chronic Toxicity

Charles *et al.*, 1996a

This review article presents data from several rat subchronic toxicity studies conducted with 2,4-D acid, 2,4-D dimethylamine salt (DMA), or 2,4-D 2-ethylhexyl ester (2-EHE). These studies were GLP Guideline studies conducted to satisfy US EPA testing requirements. The 2,4-D acid subchronic study (Schultze, 1991a) is discussed with the regulatory studies above, based on the comprehensive data in the study report. In the 2,4-D DMA and 2-EHE studies, Fischer 344 rats (10/sex/dose group) were dosed in the diet with target doses of 0, 1, 15, 100, and 300 mg/kg/day (expressed as acid equivalent doses) for 90 days.

Endocrine endpoints evaluated included: thyroid hormones (T3 and T4); adrenal, ovaries, pituitary, testes, thyroid (and parathyroids) organ weights and adrenal, epididymides, mammary gland, ovaries, pancreas, pituitary, prostate, testes, thyroid (and parathyroids), and uterus histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology, and evaluation of standard target organ histopathology were also done in these studies.

Several possible indications of endocrine pathway interactions were noted in these studies. All occurred at high and systemically toxic doses that exceeded the TSRC. These indications included decreased T4 and/or T3 at dose levels of $\geq$ 100 mg/kg/day, with T4 appearing somewhat more sensitive than T3 and females more sensitive than males. Correlating with these findings were increases in relative thyroid weights (primarily at 300 mg/kg/day); however no correlating histopathological evidence of thyroid follicular cell hypertrophy or hyperplasia was evident. Therefore, these changes are considered slight in severity and non-adverse. Adrenal cortical hypertrophy was also seen at dose levels $\geq$ 100 mg/kg/day. (Typically adrenal cortical hypertrophy results from increased ACTH release from the pituitary in response to generalized stress; in the case of 2,4-D it reflects the excessive toxicity at the high doses far exceeding the TSRC.) Relative testes weights were decreased at 300 mg/kg/day ae, and testicular atrophy was noted at the same dose level.

Treatment-related deaths were seen for 2-EHE at 300 mg/kg/day. Severe decreases in body weight gain and decreased feed consumption were seen at the high dose level (300 mg/kg/day) for all three compounds. Decreased body weight gain was also seen at 100 mg/kg/day for 2,4-D acid. A trend

towards anemia and/or decreased platelet count was observed at ≥100 mg/kg/day. BUN was increased for DMA (females) and 2-EHE at 300 mg/kg/day. Decreased relative liver and increased relative kidney weights were found at ≥100 mg/kg/day. Treatment-related histopathological findings included brush border loss and vacuolization of kidney tubular cells, centrilobular hepatocellular hypertrophy, and retinal degeneration and cataract formation (females only). These histopathological changes were primarily evident at 300 mg/kg/day.

Effects on the thyroid were considered relatively slight in severity and non-adverse, and exposure-related findings in the testes were seen only at doses clearly exceeding an MTD and far above the TSRC (and irrelevant for either human or environmental exposures). At those doses it is not clear if the observed findings reflect endocrine-related changes or are secondary to the excessive systemic toxicity. Based on other data, oxidative stress would be anticipated to occur at this very high exposure level. These studies also demonstrate that the salt and ester forms of 2,4-D tested show similar toxicities to 2,4-D when tested on an acid equivalent mg/kg/day basis.

Charles *et al.*, 1996b

This article presents data from several dog subchronic toxicity studies conducted with 2,4-D (acid), 2,4-D DMA, or 2,4-D 2-EHE, and a dog chronic toxicity study on 2,4-D acid. Data from the subchronic and chronic dog studies on 2,4-D acid (Dalgard 1993a and Dalgard 1993b, respectively) are summarized above under regulatory toxicology studies. All of these studies were GLP Guideline studies conducted to satisfy USEPA testing requirements. Beagle dogs (4/sex/dose group) were dosed in the diet with target doses of 0, 1.0, 3.75 and 7.5 mg/kg/day (expressed as acid equivalent doses.

Endocrine endpoints evaluated in these studies included adrenal, ovaries, pituitary, testes, thyroid (and parathyroids) organ weights, and adrenal, epididymides, mammary gland, ovaries, pancreas, pituitary, prostate, testes, thyroid (and parathyroids), and uterus histopathological evaluations. Clinical signs, body weight, feed consumption, clinical pathology, and evaluation of standard target organ histopathology were also performed in these studies. Thyroid hormone analyses were also conducted in the sub-chronic study on 2,4-D acid (Dalgard, 1993a).

There were two findings in these studies that potentially could be related to endocrine modulation, although it is equally possible that they are artefacts due to the young age of the study animals (4-6 months, based on the Charles *et al.* (1996b) publication. As reviewed in the discussion on the regulatory studies, testicular weight decreases and histopathological findings are exceedingly common in young dogs. Testes weights (in relation to body weight) were lower in the mid dose but not high dose of the DMA and EHE subchronic studies, and as noted above were also lower in the high dose of both subchronic studies of the acid. This finding did not occur in the chronic 2,4-D acid study (Dalgard 1993b), however, which used the same high dose as the subchronic 2,4-D acid study (Dalgard 1993a). Additionally, there were no treatment-related histopathological lesions in the testes in any of the subchronic or chronic dog studies the Dalgard 1993a and 1993b studies of the acid, or in the studies of 2,4-D DMA or 2,4-D 2-EHE. Therefore, the biological significance of the decreased testes weight finding

was considered unclear by the authors. It should be noted that an earlier subchronic study of 2,4-D acid in dogs (Schulze *et al.*, 1991b) had shown an increased incidence of testicular lesions albeit at a lethal dose (summarized above for the regulatory toxicity studies on the acid); comparison of age, body weights and the nature of the findings support that this finding was also due to the use of juvenile animals, but may have been exacerbate by systemic toxicity. The second finding in the studies reviewed by Charles *et al.* is that inactive/juvenile prostates were noted in "several high-dose males" in the subchronic studies; this finding is likely to reflect the immature age of the tested animals. Review of the report for the subchronic 2,4-D acid study (Dalgard 1993a) evaluated by Charles *et al.* shows this finding clearly present in all dose groups (see **Table S8**). This finding was also not made in the chronic acid dog toxicity study (Dalgard 1993b). As discussed above in the context of the 2,4-D acid regulatory toxicity studies, a very high incidence (up to 75%) of decreased testes weights and testicular lesions (hypospermia) have been reported in juvenile control dogs up to 9 months of age (Goedken *et al.*, 2008).

These possibly endocrine-related findings were made at doses showing generalized systemic toxicity. In the subchronic studies for all three compounds, the high-dose dogs had decreased body weight gains and decreased feed consumption; the mid-dose 2-EHE group also showed decreased weight gain. Renal toxicity was demonstrated by changes in blood urea nitrogen and creatinine, which was most pronounced at the high-dose level, and liver enzymes differed from control. No correlating histopathological lesions were found in the kidney, and only minimal increases in peri-vascular inflammation were found in the livers of high-dose dogs. Other minor clinical pathological changes were not considered toxicologically significant. None of the dog studies showed adverse effects on the thyroid.

The chronic acid study (Dalgard 1993b) showed decreased weight gain at the mid- and high dose during the first 13 weeks (correlating with the findings in the subchronic studies). No evidence of endocrine-related toxicity, including testicular toxicity, was found in this study. One to two dogs in these groups periodically showed poor weight progression and decreased feed consumption during the study.

As noted previously, dogs are not predictive for human risk assessment due to incomplete excretion of 2,4-D and lack of metabolic capabilities relevant to humans. However, as a susceptible species the dog may predict potential effects on other species deficient in the OAT-1 transporter. Dogs also provide information supporting that rodent species (especially rats) are particularly vulnerable to changes in thyroid hormone economy. Therefore data from the dog studies are included in the weight of the evidence evaluation. The publication by Charles *et al.* (1996b) also supports that findings from the dog studies on 2,4-D acid are generally predictive for those on the salts and esters of 2,4-D (when doses are expressed as acid equivalents).

Other Studies Relevant to Hypothetical Mechanism(s) for Thyroid Effects in Rats

Florsheim and Velcoff, 1962

Florsheim and Velcoff placed male Sprague-Dawley rats on a restricted iodine diet regimen for 26 days with a daily subcutaneous injection of 1 cc iodine solution (0.8 µc I131 and 5 or 10 µg stable iodide). Experimental rats were subcutaneously injected with 80 mg/kg bw/day 2,4-D for the last 7 days of the regimen. After autopsy, a 3 cc serum sample from each animal was evaluated. A significant increase in 24-hour I^{131} uptake and significant decreases in serum protein-bound-iodine (PBI) and thyroid : serum radioiodide ratio were observed. However, 2,4-D exposure had no effect on serum or pituitary TSH concentrations, thyroidal cell height, or thyroid histopathology. These findings provide a possible mechanistic explanation for decreases in circulating thyroid concentrations in rats at high dosages of 2,4-D, but do not provide evidence of a biologically significant adverse effect. Although it is not considered a high quality study, this study is cited as part of the WoE because it provides some mechanistic information to possibly explain the high-dose effects of 2,4-D on the rat thyroid.

Florsheim et al., 1963

Florsheim *et al.* (1963) evaluated the distribution of thyroxine after 2,4-D exposure. The dosage and route of administration of the 2,4-D is not specified in this paper. Based on the methodology described in Florsheim and Velcoff, it is likely that a high intraperitoneal dose was used. Rats received 2.0 milliliters (mL) of thyroxine-labeled serum by femoral infusion. No significant differences in thyroxine half-life or thyroxine distribution (30 minutes after tracer thyroxine injection) were observed. Significant differences in the fate of thyroxine 1 hour post-injection of thyroxine- I^{131} tracer were observed; however, these results were based upon data from three exposed and three control animals. Brain and liver tissues of 2,4-D treated rats had significantly higher concentrations of thyroxine- I^{131} tracer and significantly lower serum thyroxine- I131 tracer concentration compared to controls. Authors concluded that "2,4-D reduces the serum binding capacity for thyroxine..." and "2,4-D lowers serum thyroxine binding by competing with thyroxine for its binding sites, although it appears to be a weak competitor." The Florsheim *et al.* (1963) and Florsheim and Velcoff (1962) studies support that high doses of 2,4-D in the rat may modulate thyroid hormone levels. However, thyroid histopathology was not affected, even at these high doses, suggesting that the effects on circulating hormone levels had little biological significance. Although it is not considered a high quality study, this study is cited in the WoE primarily because it provides some mechanistic information to possibly explain the high-dose effects of 2,4-D on the rat thyroid.

Van den Berg, et al., 1991

Several chemical classes (including the phenoxy acids such as 2,4-D and 2,4-dichlorophenoxybutyric acid [2,4-DB]) were evaluated for their ability to bind to the T4 binding site of the thyroid hormone carrier, transthyretin. In addition, *in vivo* studies using male rats were conducted to determine the effect of selected chemicals (e.g., 2,4-DB) on plasma thyroid hormone levels. The *in vitro* study indicated that the phenoxy acids, including 2,4-D, had a relatively high affinity for the T4 binding site of transthyretin.

Because the study did not present results for individual phenoxy acids in the *in vitro* study, it cannot be determined which phenoxy compound was most reactive. The *in vivo* study showed that male rats administered 2,4-DB acutely via the intraperitoneal route had lower plasma T4 levels compared to control. 2,4-D itself was not tested in the *in vivo* study; however, the findings (for 2,4-DB) are generally consistent with the results of high-dose exposures to 2,4-D to rats in other *in vivo* studies. The intraperitoneal route of administration is not considered relevant to human risk assessment. The *in vitro* study results support that 2,4-D with its relatively high affinity for the T4 binding site of transthyretin in rats displaces the T4 from the binding protein in rats, thus making the thyroxine more available for liver sequestration or metabolism. This study is included in the WoE discussion only because it contributes a potential mechanistic hypothesis for high-dose 2,4-D effects on the rat thyroid.

Supplementary Appendix V

In vitro studies with Klimisch scores of 3 or 4[4]

General note regarding yeast based assays: In general regulatory authorities (US and EU) have moved away from supporting yeast-based assays for screening for endocrine modulating chemicals in part because of differences in absorption from mammalian systems, and also because of potential lack of stability of transfected yeast cells. Differences in rank-order potency of compounds have been detected, presumably due to differences in uptake and/or extrusion of the test compounds (Gaido *et al.*, 1997). Mammalian-based *in vitro* assays are generally considered superior to yeast, because mammalian assays are thought to be more representative of humans in terms of uptake of xenobiotics, potential biotransformation, and the cellular context of steroid receptors and the various signaling molecules with which the xenobiotics interacts (O'Connor *et al.*, 2002).

Blair *et al.*, 2000

One hundred and eighty-eight chemicals, including 2,4-D, were examined for binding at the rat estrogen receptor (ER). Sprague Dawley rats (245 days of age) were ovariectomized 10 days prior to sacrifice and the uterine tissues excised and homogenized. Chemicals were then examined for the ability to inhibit the binding of 1 nM ^{3}H-17ß-estradiol (E2) to the rat uterine ER using a tiered approach. First, chemicals were examined using two high concentrations spanning three log concentrations. Chemicals that demonstrated ER binding in Tier 1 were then examined in Tier 2 using a wider range of concentrations (typically $10^{-9} - 10^{-4}$ M). Radio-inert E2 ($10^{-10.5} - 10^{-7}$ M) was also tested as a positive reference chemical. Samples were run in duplicate and each assay was conducted twice. 2,4-D tested at concentrations up to 0.1 mM did not demonstrate an ability to bind to the rat ER, so it was not examined in a wider concentration range. In this study, the uterine tissues were derived from rats that were older than recommended in the EPA testing guidelines for the ER binding assay (EPA, 2009a). This study was conducted using the same test model as prescribed in the EPA guidelines, with the exception noted above. The results demonstrated that 2,4-D does not bind to the ER. This study was well-designed and adequately reported, however only two concentrations of 2,4-D were tested (with the second concentration undefined) and specific results for 2,4-D were not included. Therefore it was considered to score a 3 by the Klimisch criteria.

Fang *et al.*, 2003

Two hundred and two chemicals, including 2,4-D, were evaluated for the ability to bind to recombinant rat androgen receptor (AR; purchased commercially). The AR protein (1.84 nM) was incubated with 1 nM of ^{3}H-R1881 and increasing concentrations (4.28×10^{-9} M to 4.28×10^{-4} M, in one log-unit concentration intervals) of radio-inert test compounds. Radio-inert R1881 ($10^{-7} - 10^{-11}$ M) was run as a reference chemical and a no competitor tube was included to assess total radioligand binding. After 18-20 hours at 4°C, hydroxyapatite slurry was added to separate bound from unbound radioligand.

[4] See Publication **Table 4** for Klimisch scores and comments for the published *in vitro* studies.

Duplicate tubes were included at each test concentration and all assays were run at least twice. AR binding affinity was reported in terms of relative binding affinity. 2,4-D (purity not reported) failed to compete with ^{3}H-R1881 for AR binding.

The source of AR used in this study (*Escherichia coli*-expressed recombinant rat AR protein) is different from that indicated in the EPA guidelines for the AR binding assay (rat prostate; EPA, 2009b). As noted above, the purity of 2,4-D was not reported, which is the primary reason for scoring this assay as a Klimisch criteria 3. Nevertheless, the assay described in Fang *et al.* (2003) appears to be robust and to align closely with the general procedures described in the EPA guidelines, including ^{3}H-R1881 assay concentration, test compound concentrations, incubation times and temperature. It provides support for the finding that 2,4-D does not interact with the AR.

Jung *et al.*, 2004

Fifty chemicals, including 2,4-D, were evaluated for anti-estrogenic activity in a yeast two-hybrid system expressing the rat estrogen receptor ERα and the coactivator TIF2, along with a ß-galactosidase reporter gene containing an estrogen-responsive promoter region. The yeast cells were incubated overnight in SD medium free of tryptophan and leucine (required for growth), then treated with 10^{-9} - 10^{-3} M test chemicals in DMSO and 5 x 10^{-9} M E2 for 4 hours at 30°C. After incubation, the cells were washed, lysed with Zymolyse 20T, mixed with *o*-nitrophenol-ß-galactopyranoside, and galactosidase activity determined based on spectrometric absorbance for an aliquot of the lysate. Chemicals found to be exhibit anti-estrogenic activity were further evaluated using an estrogen receptor competitive binding assay and a human ER transactivation assay in MCF-7 cells. 2,4-D (form and purity not reported) was negative for anti-estrogenic activity in the yeast two-hybrid system at concentrations tested. The number of replicates run, the number of chemical concentrations tested, and data regarding solubility of the test chemicals and cytotoxicity were not provided. Another limitation of the Jung *et al.* paper involves the use of a yeast-based assay system (see general note above).

Jungbauer and Beck, 2002

Various compounds, including 2,4-D (form and purity not reported), were tested for estrogenic activity in a yeast two-hybrid system expressing the human ERα (the expression of which was under control of a metallothionine promoter) and a reporter plasmid containing the lacZ gene under control of the vitellogenin hormone response element. Human ER expression was induced overnight by the addition of 10^{-5} M $CuSO_4$. Cells were then treated with test chemicals for 4 hours at 30°C, after which the induced expression of ß-galactosidase was determined spectrophotometrically. E2 was run as a reference chemical to characterize the system. Potency values within the range of 10^{-4} M were considered not biologically relevant. Exact details regarding assay parameters were limited (e.g., concentrations of test chemicals evaluated, number of replicates and runs of the assay). 2,4-D did not transactivate the estrogen receptor. Another limitation of the Jungbauer and Beck paper involves the use of a yeast-based assay system (see general note above).

Kim *et al.*, 2005

This paper contained the results of several different *in vitro* assays. One concern potentially limiting relevance is that the 10nM concentration of DHT used in most of the assays is potentially a supra-physiological concentration, based on serum and prostate DHT concentrations reported in Belanger et al., 1989.

22Rv1 cell proliferation

This assay evaluated the ability of 2,4-D (98% purity) and its metabolite, 2-4-dichlorophenol (DCP; 98% purity), to alter the response of human prostate cancer cells to dihydroxytestosterone (DHT) treatment. Human 22Rv1 cells, which express the AR, were seeded in 96-well plates at 3×10^{-3} cells per well in RPMI medium containing 10% fetal bovine serum (FBS) pre-treated with dextran-coated charcoal. Twenty-four hours later, the cells were treated with $10^{-13} - 10^{-5}$ M 2,4-D, DCP, or DHT or with $10^{-9} - 10^{-5}$ M 2,4-D or DCP in the presence of 10^{-8} M DHT. After four days, the cells were fixed and stained with sulforhodamine-B to examine cell proliferation spectrophotometrically. Reported results were the mean of three separate experiments. Neither 2,4-D nor DCP alone increased 22Rv1 cell proliferation; however, co-treatment combining either 2,4-D or DCP and DHT was reported to increase cell proliferation substantially above levels induced by DHT treatment alone. No solvent only controls were tested in this assay.

Note in a 2006 abstract by the same group of investigators (not relating to 2,4-D), DHT alone was reported not to have induced proliferation in this cell line, in contrast to what was reported in the 2005 study.

Transactivation in 22Rv1 cells and PC3 cells

Similar results were reported in receptor transactivation experiments. 22Rv1 cells were transiently transfected with a mouse mammary tumor virus-linked luciferase reporter gene, then treated 24 hours later with 10^{-8} M DHT, $10^{-9} - 10^{-5}$ M 2,4-D or DCP, or 10^{-8} M DHT plus $10^{-12} - 10^{-6}$ M 2,4-D or DCP. Likewise, PC3 cells, which do not express the AR, were transiently transfected with the luciferase reporter gene as well as a human AR expression vector, then treated 24 hours later with 10^{-8} M DHT, 2,4-D, DCP, or a combination of both DHT plus 2,4-D or DCP. After 24 hours incubation, cells were lysed and extracts assayed for luciferase activity, which was normalized for transfection efficiency using expression of a Tk vector. Neither 2,4-D or DCP alone caused transactivation of the AR in 22Rv1 or PC3 cells; however, both chemicals were reported to substantially increase the AR transactivation induced by treatment with 10^{-8} M DHT. No solvent only controls were tested in this assay.

AR mRNA Expression

Treatment of 22Rv1 cells with 10^{-8} M DHT, 10^{-7} M 2,4-D, 10^{-10} M DCP, or a combination of DHT with 2,4-D or DCP had no significant effect on AR mRNA expression (as assessed by slot blot analysis with normalization against ß-actin mRNA expression). Although DHT treatment induced AR protein expression (as assessed by Western blot analysis with normalization against ß-actin protein expression)

approximately 50%, treatment with 2,4-D, DCP had no effect. Additionally, co-treatment with DHT and 2,4-D and DCP did not further increase AR protein expression above levels induced by DHT treatment alone. No solvent only controls were tested in this assay and only one concentration of 2,4-D was evaluated

AR Receptor Binding

The ability of 2,4-D and DCP to bind to the AR receptor was evaluated in COS-1 cells (African green monkey kidney cells) transiently transfected with the human AR. Following transfection, cells were cultured in serum-free and phenol-free DMEM for 24 hours, then treated for 2 hours with 5×10^{-9} M ^{3}H-DHT in the presence or absence of 5×10^{-9} M, 5×10^{-8} M, or 5×10^{-7} M of DHT, 2,4-D or DCP. Both 2,4-D and DCP inhibited binding of ^{3}H-DHT by approximately 50% at a concentration of 5×10^{-7} M. No solvent only controls were tested in this assay.

AR Nuclear Translocation

Finally, the ability of 2,4-D and DCP to affect AR nuclear translocation was investigated in PC3 cells that were transiently transfected with a green fluorescent protein (GFP)-AR fusion protein expression vector. Cells were treated for 20 min at 37°C to 10^{-8} M DHT, 2,4-D, or DCP or 10^{-8} M DHT plus 10^{-8} M 2,4-D or DCP. After incubation, cells were fixed in 10% formaldehyde and examined for fluorescence using a fluorescence microscope. Treatment with DHT, but not 2,4-D or DCP, induced nuclear translocation of the GFP-AR fusion protein. Co-treatment with DHT plus 2,4-D or DCP was reported to result in nuclear translocation that was increased over that induced by DHT treatment alone. No solvent only controls were tested in this assay.

The authors of this study interpreted the results to suggest that 2,4-D and its environmental metabolite, DCP, do not activate the human AR receptor; but they hypothesized that these compounds may enhance the activity of DHT at the AR by promoting nuclear translocation of the DHT-activated receptor. Kim *et al.* (2005) used a very different, and less robust, model system (COS-1 cells transiently transfected with human AR) than that indicated in EPA guidelines for the assessment of AR binding (2009b). For one, they evaluated binding in whole cells rather than using receptor protein from homogenized tissues or lysed cells. As well, they did not use a reagent such as hydroxyapatite to separate unbound from bound ligand, only evaluated a limited number of test compound concentrations compared to the range recommended in the EPA testing guidelines, and did not report using a solvent/vehicle only control in this assay, or any of the others described in the paper. The findings in Kim *et al.* contradict those of Fang *et al.* (2003), which used recombinant rat AR protein and tested a wider range of test compound concentrations (4.28×10^{-9} to 4.28×10^{-4} M) although the Fang et al. assay did not report the purity of 2,4-D tested. The results also contradict the results of the ToxCast™ program screening both in cell-free and cell-based assays. Finally, although 2,4-D and DCP were found to bind the human AR in the study of Kim *et al.* (2005), these compounds alone did not induce proliferation of an androgen-response prostate cancer cell line or activate the human AR, as demonstrated in the receptor transactivation assays. This study is considered limited and the findings are difficult to interpret.

Lee *et al.*, 2006

Lee *et al.* (2006) developed a yeast two-hybrid detection system to examine transactivation via the estrogenic receptor for a variety of compounds. In this experiment, yeast was created that expressed bait, fish and reporter gene. The "bait" were human estrogen receptor α (ER α), either full length or with the transactivation domains missing (truncated), fused to a GAL4 DNA binding domain. This chimeric ER α (full length or truncated) would bind to DNA as a monomer without the need for a ligand. The "fish" were steroid receptor co-activator-1 (SRC-1) or transcriptional intermediate factor-2 (TIF-2) that was fused to the GAL4 transactivation domain. The reporter gene consisted of the GAL4 DNA binding domain controlling a *LacZ* gene. Agonist activation of the DNA bound bait (ER α – Gal4) allows the fish (SRC-1 or TIF-2) to bind to the bait and this induces the reporter gene. This artificial system is quite different than normal mammalian estrogen signaling where agonist activated ER α first binds to another ER α protein forming a ER α homodimer that only then would bind to an ER DNA binding site (ERE) and activate a target gene.

The authors noted increased galactosidase activity at lower concentrations of the test compounds than the original one-hybrid construct. 2,4-D, which was negative with the original one-hybrid construct, had detectable binding with the two-hybrid construct, ranging from 2.09×10^{-4} M to 5.42×10^{-6} M (17β-estradiol was positive at 10^{-10}M).

There were no negative control compounds included in the Lee *et al.* (2006) study. Another limitation of the Lee *et al.* (2006) paper involves the use of a yeast-based assay system (see note above). Another issue is that agonists and antagonists do not always mimic their *in vivo* agonist or antagonist activity (Pham *et al.*, 1991, 1992). It is possible that some of this discrepancy is due to the use of an artificial system in which the conformational shapes of the receptors and differences in post-translational modifications across species alter their response pattern to exogenous compounds (Mak *et al.*, 1989; Pham *et al.*, 1992; Routledge and Sumpter, 1996; Zysk *et al.*, 1995). Thus, this study is considered limited and the findings are difficult to interpret.

Lemaire *et al.*, 2006

Forty-nine pesticides, including 2,4-D (purity of 95% or greater), were tested for the ability to activate the human ERα and ERβ using stably transfected HeLa cells expressing either ERα or ERβ and luciferase reporter gene with an upstream estrogen responsive element. Cells were seeded at a density of 30,000 cells per well in phenol red-free DMEM supplemented with 6% dextran-coated charcoal fetal calf serum. After 24 hours, cells were incubated with 10 µM test chemical for 16 hours, after which luciferase activity was measured using a luminometer. ERα and ERβ receptor activity was reported as a percentage of the activity induced by 10 nM 17ß-estradiol and findings are the result of three separate experiments conducted with either duplicate or triplicate wells for each transfected cell line. Chemicals were initially tested at a single concentration of 10 µM; testing at lower concentrations was conducted only if a positive response was seen at the high initial concentration. The transfected cell lines also expressed a constitutive background level of luciferase activity and the ability of the test chemicals to affect that activity was also evaluated. At a concentration of 10 µM, 2,4-D exhibited no differences from

a DMSO control for relative activity to both the ERα and ERβ. Further, it did not affect the constitutive activity of either receptor protein in the absence of 17ß-estradiol. 2,4-D was not further tested for estrogen receptor binding in this study.

This study evaluates antagonism activity of constitutive ERα activity and the absence of both agonist and antagonist activity at ERß. Although the results were clearly negative, 2,4-D was tested in only a single high concentration; this is the major weakness identified in this study.

Nishihara *et al.*, 2000

In this study, 517 chemicals, including 2,4-D, were tested for agonist activity at the estrogen receptor in a yeast two-hybrid system expressing the estrogen receptor ERα and the coactivator TIF2, along with a ß-galactosidase reporter gene containing an estrogen-responsive promoter region. The yeast cells were incubated overnight in SD medium free of tryptophan and leucine (required for growth), then treated with test chemicals in DMSO for 4 hours at 30°C. After incubation, the cells were washed, lysed with Zymolyse 20T, mixed with *o*-nitrophenol-ß-galactoside, and galactosidase activity determined based on spectrometric absorbance for an aliquot of the lysate. 2,4-D (form and purity not reported) was negative for estrogenic activity at concentrations up to 1×10^{-4} M. The number of replicates run, the number of chemical concentrations tested, and data regarding solubility of the test chemicals and cytotoxicity were not provided. Another limitation of the Nishihara *et al.* (2000) assay involves the use of a yeast-based assay system (see note above).

Orton *et al.*, 2009

Twelve pesticides, including 2,4-D (>97% purity), were tested for agonist and antagonist activity at both the human estrogen receptor and the human androgen receptor using yeast reporter systems. The DNA sequence of either the human estrogen receptor (yeast estrogen screen; YES) or the human androgen receptor (yeast androgen screen; YAS) was stably integrated into the yeast genome. The yeast cells also contained a lac-Z reporter plasmid, which encodes the ß-galactosidase enzyme and could be expressed upon activation of either the estrogen or androgen receptor. For assessing agonist activity, concentrations of $4.9 \times 10^{-7} - 1 \times 10^{-3}$ M 2,4-D dissolved in ethanol were added to the wells of a 96-well plate and evaporated to dryness. Aliquots (200 µl) of minimal medium seeded with either the YES or YAS yeast cells and containing the chromogenic substrate chlorophenol red-ß-D-galactopyranoside (CPRG) were added to the wells and incubated at 32°C for 3 days. Samples were run in triplicate over three plates and the assay was run two separate times. For the YES assay, 17ß-estradiol was run as a positive control; for the YAS assay, testosterone was run as a positive control. Both solvent and media-only controls were also run. Conversion of the yellow CPRG substrate to a red product was then assessed by measuring absorbance at 540 nm. Cell growth (based on culture turbidity measured at 620 nm) was also measured to assess cytotoxicity. To evaluate antagonist activity, the assays were run similarly, except the pesticide was added in combination with either 0.25 nM 17ß-estradiol (for the YES system) or 2.5 nM testosterone (for the YAS system). Positive controls were either 0.01-25 µM 4-hydroxytamoxifen (to assess anti-estrogenic activity) or 0.02-50 µM flutamide (to assess anti-androgenic activity). In these assays, 2,4-D was stated to be negative for agonist and antagonist activity at both the

human estrogen receptor and human androgen receptor. No 2,4-D specific data were presented, however.

The YES and YAS assays were well run with appropriate positive controls and with both negative and vehicle controls. The primary limitation of this assay reported by Orton *et al.* involves the use of a yeast-based assay system, as discussed in the note above. The lack of specific data for 2,4-D is also considered a weakness.

In addition to the yeast assays discussed above, Orton *et al.* also evaluated the effects of pesticide exposure on the frog ovarian production of progesterone, testosterone and estradiol *in vitro* and on the number of oocytes ovulated during the exposure period. Ovaries were removed from sexually mature female *Xenopus laevis*, sliced into 10 fragments, and cultured in Modified Barth's Media in 24-well plates (2 fragments per well). After 20 hrs incubation (incubation temperature not reported) with varying concentrations of human chorionic gonadotropin (hCG) or media only, fragments were incubated for 20 hrs in either control media, a concentration of hCG causing a 60% maximum ovulatory response, or 6.25 or 62.5 µM pesticide plus hCG. Samples of media were taken after both the initial and subsequent incubation periods for evaluation of progesterone, testosterone, and estradiol concentrations via radioimmunoassay. After incubation, oocytes were tested for viability using trypan blue and fixed with trichloroacetic acid. The number of ovulated oocytes (determined via dissociation from the follicular tissues and the presence of a "white spot") were counted. In each assay run, six replicate samples were examined for each treatment and the assay was run three separate times.

There are several reservations regarding the performance of this assay. The stage of the collected oocytes used by Orton *et al.* appears fairly random, although a higher proportion of stage V-VI ovaries were collected (for the 10 fragments/ovary, 7 or 8 were stage V/VI and 2 or 3 stage I-IV). Two fragments were used for each test well, but it does not appear that distribution of fragments to the wells was systematic regarding the stages distributed. Sretarugsa and Wallace (1997) report considerable differences for steroid hormone production from frog oocytes at different stages, with optimum steroidogenesis at stages IV and VI but a marked shift in the ratio of progesterone to estradiol production over this development window. This suggests that chance assortment of relatively more or less mature oocytes could have significantly altered the assay outcomes. Further, only two concentrations of 2,4-D were tested.

2,4-D treatment had no effect on ovulation or on steroidogenesis under the conditions of this assay; but the reservations regarding the assay methods should be noted.

Petit *et al.*, 1997

Forty-nine chemicals, including 2,4-D, were evaluated for estrogenic activity in three different *in vitro* assays. The ability of test compounds to induce vitellogenin mRNA expression was evaluated using primary hepatocyte cultures derived from male rainbow trout (250-750 grams each). Liver cells were harvested and 1-2 x 10^7 cells per petri dish cultured in DMEM/Ham's F-12 (without phenol red) plus additional supplements. Once aggregates formed, cells were treated for 48 hours with either E2 (10^{-6}

M) or test compound (10^{-4} M or concentration that gave maximal ß-galactosidase activity in yeast system); mRNA was then harvested and vitellogenin mRNA expression assessed by slot blot assay (normalizing against ß-actin expression). Reported results represent the mean values of 3 to 16 independent experiments. In the case of 2,4-D, vitellogenin expression was slightly greater than that induced by solvent alone; however, the study authors felt that the level of induction was too close to control levels to determine if it was ineffective or weakly estrogenic. Weaknesses include that only a single concentration of 2,4-D was tested.

ER transactivation was evaluated using a stably transfected yeast system expressing both the rainbow trout ER and a ß-galactosidase reporter gene downstream of two tandem copies of the consensus ERE. Yeast cells in liquid culture were incubated for 4 hours at 30°C to $10^{-8} - 10^{-4}$ M of test chemicals or E2, then assayed for ß-galactosidase activity using a spectrophotometer. Reported results were based on mean values from at least three separate experiments. Treatment with 2,4-D (purity and form not reported) did not increase ß-galactosidase activity significantly above solvent control levels, indicating the absence of rainbow trout ER transactivation.

Rainbow trout ER binding was assessed using the same yeast system described above. Whole yeast extracts were incubated for 16 hours at 4°C to 20 nM ^{3}H-E2 in the presence or absence of increasing concentrations ($2 \times 10^{-6} - 2 \times 10^{-3}$ M) of radio-inert test compounds, after which samples were treated with dextran-coated charcoal to separate free from bound ^{3}H-E2. Radio-inert E2 ($2 \times 10^{-9} - 2 \times 10^{-5}$ M) was run as a positive reference chemical. Results were based on mean values from duplicate samples tested at each concentration. 2,4-D did not competitively bind to the rainbow trout ER. The primary limitation of the second and third assays reported by Petit *et al.* involves the use of a yeast-based assay system, as discussed in the note above.

Soto *et al.*, 1995

Various chemicals, including 2,4-D, were examined for estrogenicity using the E-SCREEN assay. MCF-7 cells were seeded into DMEM supplemented with 5% FBS at a density of 2×10^{5} cells per well. After 24 hours, the medium was removed, replaced with phenol red-free DMEM supplemented with 5% charcoal-dextran treated human serum, and a range of test compound concentrations added (exact concentrations not reported). After six days of incubation, the cells were lysed and nuclei counted as a measure of cell number. Both the relative proliferative potency (ratio between minimal concentration of E2 need for maximum cell yield and the minimal dose of test compound for a similar effect) and the relative proliferative effect (100 times the ratio between the highest cell yield obtained with the chemical and with E2) were determined for each chemical. Reported results were the mean of at least five separate experiments using duplicate samples in each. In this study, 2,4-D (purity not reported) was considered non-estrogenic. Additional studies of ER and progesterone receptor (PR) processing, induction of pS2 expression (an estrogen responsive gene), and binding to the ER were conducted; however, these studies appear to have only been done on those compounds that were positive in the E-SCREEN, hence 2,4-D was not tested in these additional assays.

Vonier *et al.*, 1996

Various pesticides, including 2,4-D, were evaluated for the ability to competitively bind to the ER and progesterone receptor (PR) extracted from oviduct tissue of adult female alligators (*Alligator mississippiensis*) captured in central Florida. Frozen tissues were thawed on ice, homogenized in buffer, centrifuged, and supernatants taken for subsequent receptor binding assays. To assess ER binding, protein extracts were incubated with 2.5 nM ^{3}H-E2 for 1 hour at 25°C, after which any free ^{3}H-E2 was removed using a charcoal-dextran mixture. Samples were then incubated with increasing concentrations of radio-inert test chemicals (exact concentrations not reported) for 1 hour at 25°C. Free ^{3}H-E2 was again removed using the charcoal-dextran mixture and samples analyzed for ^{3}H-E2 binding. To evaluate PR binding, samples were incubated at 4C for 12 hours with ^{3}H-R5020 in the presence or absence of increasing concentrations of radio-inert test chemicals. Free ^{3}H-R5020 was removed using a charcoal-dextran mixture and samples analyzed for ^{3}H-R5020 binding. Findings for both receptor binding assays are the mean results of at least three independent experiments using three replicates each. 2,4-D did not bind the ER or PR to any significant extent in these experiments.

This study was conducted using a different model system (alligator ER) than that specified in the EPA guidelines for the ER binding assay (EPA, 2009a) and the exact range of test compound concentrations examined were not reported (Klimisch 3). Nevertheless, this study provides additional support indicating that 2,4-D does not bind to the ER. Additionally, it shows that 2,4-D does not act at the PR either.

Supplementary Appendix VI

Ecotoxicological Studies with Klimisch Score of 3 or 4[5]

Amphibians

Aronzon *et al.* (2011)

Aronzon *et al.* evaluated acute and short term toxicity of 2,4-D DBE (99% purity) and a commercial formulation (CF) to South American Toad, *Rhinella arenarum*, embryos. Assays were conducted with the AMPHIEMB protocol for early life stage toxicity and AMPHISHORT for seven-day toxicity. Duplicate batches of 10 embryos were treated with either 2,4-D DBE or with CF. Experiments were repeated four times.

Embryos were exposed

(i) continuously from 2–4 blastomere stage up to the end of embryonic development in concentrations ranging from 1 to 15 mg/L 2,4-D DBE and from 0.25 to 3.85 mg/L of CF,

[5] See Publication **Table 7** for Klimisch scores and comments for published ecotoxicological studies

(ii) in pulses during the blastula, gastrula, tail bud, muscular activity, gill circulation, open mouth, opercular folds stages for 24 hours with doses ranging from 1 to 26 mg/L for 2,4-D DBE and 0.1 to 10 mg/L for CF, and

(iii) continuously from the end of embryonic development for 168 hours to concentrations ranging from 8 to 20 mg/L for 2,4-D DBE and 2 to 5 mg/L for CF.

Adverse effects were analyzed for lethality, malformations, stage-dependent susceptibility, and ultrastructural features. The CF was significantly more toxic than 2,4-D DBE in all cases where embryos were continuously exposed from the blastula stage onwards (2.68 ± 0.05), in all pulse treatment experiments, and in post-embryonic development continuous exposure studies (4.45 ± 0.37). The authors stated both compounds were teratogenic. However, it appears no controls were run for these experiments and therefore it is not possible to interpret the results.

Heggstrom, 2009

Wood frog tadpoles (*Rana sylvatica*) were exposed to nominal concentrations of 0.1, 1.0, and 100 µg/L 2,4-D dimethylamine (99% pure) in microcosms from shortly after the time of hatch until metamorphic climax (approximately 30- 56 days depending on tadpole development). Forty tadpoles were placed in each microcosm and five replicate microcosms were used per treatment level. Survival, deformities, time to metamorphic climax and morphometric measures including total length, snout-vent length, and wet weight were analyzed. In addition, circulating levels of cortisol were determined from successful blood collections. There were no 2,4-D dimethylamine treatment-related effects on any of the measured parameters, except for total length, which was observed to be significantly decreased in the 0.1 µg/L treatment group (but not the 1.0 or 100 µg/L treatment groups) relative to controls. The biological significance of this finding is unknown, since this endpoint was reported to be highly variable at metamorphic completion. The absence of dose response suggests this is not an exposure-related effect. Additionally, high mortality among the tadpoles was observed in the microcosms across all treatment groups, possibly due to viral infections or latex toxicity (due to lab gloves), thus the interpretation of this microcosm study in regard to 2,4-D dimethylamine toxicity is problematic.

The same basic experimental design and measurement endpoints were used with wood frog tadpoles in field ponds (forested, agricultural, agricultural + 10 µg/L 2,4-D dimethylamine application). The results from the field pond studies indicated that tadpoles in the agricultural ponds were larger than the tadpoles in the forested ponds. Additionally, tadpoles in the agricultural ponds had increased cortisol levels compared to tadpoles in the forested pond. The relevance of 2,4-D exposure on the observed endpoints cannot be ascertained from the field pond study since tadpoles from the agricultural pond in the absence of 2,4-D dimethylamine and tadpoles from the agricultural pond applied with 10 µg/L 2,4-D dimethylamine gave similar results in the measured endpoints.

LaChapelle *et al.*, 2007

Xenopus oocytes were harvested from anesthetized adult female frogs and exposed to 2,4-D (purity not reported) for 48 hours to determine the mechanism of 2,4-D inhibition of oocyte maturation. The exposure concentration was 2.43 g/L (10 mM). Oocytes were microinjected with radiolabelled *in vitro*

transcripts of Mos RNA and exposed to progesterone and 2,4-D to decipher the mechanism of 2,4-D dysfunction of the meiotic signaling mechanism. 2,4-D induced irreversible dysfunction of the meiotic signaling mechanism and this was linked to inhibition in the cascade of signaling events from membrane-bound progesterone receptor binding. The investigators indicated the results show that 2,4-D may inhibit progesterone induced gamete maturation. This study was a mechanistic study and tested only a single extremely high concentration of 2,4-D. The concentration of 2,4-D evaluated in this study was orders of magnitude greater (approximately 1,000 x greater) than an environmentally relevant dose. Furthermore, a dose of this level would be expected to result in overt toxicity (lethality) *in vivo*, based on the LC50 for 2,4-D ranging from 225 to 359 mg/L among various species of frogs (Palmer and Krueger, 1997; Morgan et al., 1996). It is not considered relevant to any realistic exposure scenario.

Morgan *et al.*, 1996

The teratogenic potential of a "commercial formulation" of 2,4-D (Aldrich Chemical Company, 99% pure) was evaluated in the frog embryo teratogenic assay-Xenopus (FETAX). The FETAX test is a 96-hour static renewal bioassay used to determine the teratogenic potential of chemicals. Toxicity of 2,4-D toward *Xenopus* embryos was assessed in FETAX assay buffer and in natural water. The natural water used in this study is of unknown composition, and may be lacking essential ingredients for normal development, thus the results of the evaluation in FETAX assay buffer (supplemented with essential cations and anions for proper embryo development) is discussed. Fertilized eggs were harvested from breeding adult *Xenopus laevis* frogs and de-jellied. The de-jellied embryos were placed in 100 mm x 15 mm plastic petri dishes (20 embryos/plate and four replicate plates per treatment) and exposed to waterborne concentrations of 2,4-D for 96 hours with daily renewals of test solutions. Nominal 2,4-D concentrations were 180, 190, 210, 220, 230, 240, 250, 260, and 270 mg/L. The LOEC, EC50 (malformations), and LC50 for 2,4-D were determined to be 226, 245, and 254 mg/L, respectively. Both mild and severe edema of the gut were noted among tadpoles exposed to 2,4-D. The results of this study indicate that 2,4-D is teratogenic to *Xenopus laevis* embryos only at high concentrations, similar to those concentrations causing lethality. This study is not considered useful for the weight-of-the evidence evaluation.

Stebbins-Boaz *et al.*, 2004

Xenopus oocytes were harvested from anesthetized adult female frogs and exposed in vitro to 2.5 - 10 mM 2,4-D sodium salt (0.6 – 2.43 g/L) at pH 6.8 for variable periods of time (1-10 hours). At the concentrations of 2,4-D evaluated, germinal vesicle breakdown (GVBD) was blocked in the presence of progesterone (via inhibition of the Mos protein via blocking post translational polyadenylation of mRNA) and also microtubule depolymerization was observed.. The results suggested that 2,4-D requires microfilaments (actin) to initiate observed alterations in oocyte morphology. Also, 2,4-D induced MAPK activation. There was no discussion of the environmental relevance of these findings; rather the paper was mechanistic in focus. The concentrations of 2,4-D used in this *in vitro* study were orders of magnitude above expected environmental concentrations and were also above concentrations that are associated with acute lethality in various species of frogs. It is not considered useful for the weight of evidence evaluation.

Vardia *et al.*, 1984

Indian toad tadpoles (*Bufo melanostictus*) were exposed to waterborne concentrations of 2,4-D (source unknown; assumed to be commercially available formulation) over 96 hours in a static renewal exposure system. The concentrations of 2,4-D used were 0, 7.5, 10 and 11 mg/L. One liter beakers were used as test vessels. The temperature during experimentation was around 25 °C, the pH was around 8.3, and total alkalinity and hardness were 210 and 220 mg/L CaCO3, respectively. The 96-hour LC50 value for 2,4-D was reported to be 8.05 mg/L. Little experimental detail is discussed in this publication and the actual form of 2,4-D tested is unknown. This study lacks sufficient information to be useful for assessment of potential endocrine effects.

Fish

Xie *et al.*, 2005

In this investigation juvenile rainbow trout (*Oncorhynchus mykiss*) of unknown sex (standard length 11.5 ± 2.2 cm) were exposed to 2,4-D dimethylamine (Nufarm Co St. Joseph, MO, USA; purity not stated) for 7 days in a daily static renewal system. Several different exposure scenarios with 2,4-D dimethylamine were conducted. Rainbow trout were exposed to one "worst case" concentration of 1.64 mg/L 2,4-D dimethylamine (measured), and also to 2,4-D dimethylamine at concentrations of 0, 0.00164, 0.0164, 0.164, and 1.64 mg/L). Rainbow trout were exposed to 2,4-D dimethylamine (at the same nominal concentrations) in combination with the surfactants R-11 and Target Prospreader Activator (TPA). Each test concentration level was composed of three replicate tanks with two fish in each tank, in a static test system.

Following exposure, blood was collected from each fish, and vitellogenin levels were measured in the blood plasma with the use of a commercially available rainbow trout ELISA kit (Biosense, Bergen, Norway). In the first exposure to the "worst case" concentration of 2,4-D (1.64 mg/L), vitellogenin levels were significantly greater than control levels. In combination with the surfactant R-11, the same concentration of 2,4-D, however, did not result in vitellogenin levels that were significantly greater than controls. In combination with the surfactant TPA, however, the same concentration of 2,4-D did result in vitellogenin levels that were significantly greater than controls. In the dose-response study with 2,4-D, concentrations ≥ 0.164 mg/L resulted in significantly greater concentrations of vitellogenin compared to controls. Thus the NOEC and LOEC for vitellogenin induction in the rainbow trout were reported as 0.0164 and 0.164 mg/L, respectively. These values were altered based on co-treatment with both R11 and TPA surfactants. The NOEC and LOEC for 2,4-D in the presence of various concentrations of TPA were lowered to 0.00164 and 0.0164 mg/L, respectively. The NOEC and LOEC for 2,4-D in the presence of various concentrations of R-11 were increased to 0.164 and 1.64 mg/L, respectively.

The sample size is small, especially for measuring inherently variable endocrine parameters such as vitellogenin concentrations. Although not specifically reported, the variability around these measured vitellogenin concentrations was notably high (by visual inspection of the figures). For example, control vitellogenin responses ranged from near zero to as high as approximately 10 ng/mg total plasma protein; such variation confounds clear identification of NOEC and LOEC values reported in this study. Xie *et al.* also noted that a combined treatment with 1.64 mg/L 2,4-D and 1.46 mg/L of the surfactant R-

11 resulted in approximately 16-18 ng/mg vitellogenin expression. A subsequent experiment, however, using a similar treatment period and 2,4-D concentration (1.64 mg/L) but a lower concentration of R-11 (0.89 mg/L) resulted in much higher vitellogenin expression (approximately 50 ng/mg).

The sex and age of the juvenile fish were not reported in this study. Schlenck et al (2008) in a rebuttal to an unpublished critique (Kramer *et al.*) on the Xie *et al.* study indicated that the fish were young juveniles and sex would therefore not influence vitellogenin levels. However, detectable VTG has been reported in immature female trout (Bon *et al.*, 1997). It is possible therefore that endogenous VTG in any female fish confounds interpretation of the study results.

No effects on vitellogenin were observed in the 2,4-D FSTR assay (Coady *et al.*, 2013), in contrast to the reported findings by Xie *et al.* This may be due to species differences (the FSTRA was in fathead minnows, versus the Xie *et al.* study in rainbow trout). Schlenk *et al.* (2008) postulated the observed effects in the Xie *et al.* study might be due to a 2,4-D metabolite; this does not however explain the absence of effects on vitellogenin in the Coady *et al.* (2013) FSTRA study which exposed more fish to 2,4-D for a longer duration.

As the Xie *et al.* test system was static it is also possible 2,4-dichlorophenol (2,4-DCP) accumulated as an aquatic 2,4-D degradate (Coady *et al.* used a flow through system). 2,4-DCP is a photolysis/hydrolysis environmental metabolite of 2,4-D, and exhibits relatively weak in *vitro* estrogen receptor binding activity (Li *et al.*, 2012; Nishihara *et al.*, 2000). Ma *et al.* (2012), reported adverse effects on zebrafish reproduction in fish exposed to 2,4-DCP at 0.3 mg/L; it is unclear whether these effects are causally linked to an estrogenic mode of action, or whether decreased reproduction is related to general systemic toxicity, since the acute toxicity for 2,4-DCP exposed zebrafish is only approximately 10-fold greater (LC50 of 3.9 mg/L (ECOTOX database)). Ma *et al.* (2012) concluded that there was potential evidence of changes in steroidogenesis that might affect reproduction. The Ma *et al.* study appeared to be reasonably well conducted, although hormonal and reproductive data were variable and the concentrations of 2,4-DCP evaluated were above what would be expected in the environment.

Thus, the relatively small sample size, variable nature of the reported responses and unknown age and sex of the fish studied by Xie *et al.* (2005) requires further investigation and replication before inferring either fish or mammalian environmental health risks.

Holcombe *et al.*, 1995

Japanese medaka (*Oryzias latipes*) were exposed to 2,4-D acid (CAS # 94757; 99% pure) in 96- hour acute and two 28-day larval survival and growth tests. In the chronic test, exposure to 2,4-D acid was discontinued after 28 days at which time live fish weights were taken and fish were then transferred to clean water for an additional 5 months. At the end of a five month period, fish tissues were assessed pathologically for incidences of tumors (data not reported in this publication). 2,4-D stock solutions were pH adjusted with NaOH and HNO3. A continuous-flow mini diluter exposure system with a flow rate of 25 ml/min was used for both the acute and chronic tests. Water chemistry data including dissolved oxygen, pH, temperature, hardness, and alkalinity were measured during the exposure and were within acceptable ranges for medaka. The medaka used for all tests were obtained from the

Environmental Research Laboratory-Duluth culture unit. Twenty medaka (10 per replicate) were exposed per concentration in the acute tests. Mean measured concentrations of 2,4-D acid in the acute study were < detection limit (control), 567, 1100, 2250, 4560, and 8970 mg/L. The 96-hour LC50 for medaka was determined to be 2780 mg/L 2,4-D Acid. Sixty medaka larvae were stocked in each test vessel (18.5 x 14.0 x 13.0 cm deep) in the chronic studies. Two replicate vessels per test concentration were used in both of the 28-day chronic toxicity tests. Mean measured concentrations of 2,4-D acid in the first 28-day study were < detection limit (control), 27.2, 56.5, 113, 221, 425 mg/L. Mean measured concentrations of 2,4-D acid in the second 28-day study were < detection limit (control), 2.37, 5.73, 13.5, 30, 60.2 mg/L. Survival and growth of medaka were both significantly reduced by 2,4-D concentrations of 56.5 mg/L and 60.2 mg/L during test #1 and test #2, respectively. Although the study quality is adequate, there are no specific endpoints related to endocrine activity in this study.

Koc and Akbulut, 2012

Koc and Akbulut exposed zebrafish for 5 days to 2,4-D (purity not defined, possibly a formulation) at concentrations of 0.1-1 mg/L. No dose analyses were done. They reported ovarian histopathological changes and atretic follicles with a dose-related increase in severity. Selected slides of ovarian tissue were presented. The control slide was at a different level of magnification than slides from exposure fish, making direct comparison difficult. Additionally, some of the slides show different staining (suggesting excessive heat during tissue processing), and sectioning artifacts (characteristic of a dull microtome blade). Despite these problems there were some marked differences in the appearance of follicles between the control and exposed groups. In the absence of information regarding the test material and in view of possible processing and sectioning artifacts in the ovarian histology, this study is not considered of sufficient quality to include in the WoE.

Padilla *et al.*, 2012

As part of the Computational Toxicology Research Program of the U.S. EPA, the toxicity of the 309 ToxCast™ Phase I chemicals was assessed using a zebrafish screen for developmental toxicity. Exposure (immersion) was from 6–8 hours post fertilization to 5 days post fertilization. 2,4-D (purity greater than 90%) in aqueous DMSO was tested in a single concentration study at a relatively high nominal dose (80 µM). Five days post fertilization larvae were assessed for death and overt structural defects. 2,4-D was reported as "positive" in the single high concentration study, although the degree of response was weak. Further details of the response were not specified. 2,4-D was then tested in a multiple concentration study, with concentrations ranging from 0.001 to 80 µM. 2,4-D was negative in this study and no AC50 was calculated. Although this study was considered relatively reliable (Klimisch score of 2) the findings reported in this study are too non-specific to provide evidence of potential endocrine pathway interactions (and further were not replicated).

Rehwold *et al.*, 1977

Acute toxicity tests (96 hour) with 2,4-D (no form or purity information given) were conducted with striped bass (*Marone saxatillis*), the banded killifish (*Fundulus diaphanous*), pumpkinseed, white perch (*Roccus americanus*), American eel (*Anguilla rostrata*), carp (*Cyprinus carpio*) and guppy (*Libistes*

reticulatus). All fish species, except for the guppies, were collected from the Hudson River. Guppies were purchased from a pet store. 2,4-D was obtained from Analabs, Inc. The experimental design and test concentrations were not described in detail in the publication. 96-hour LC50 values ranged from 26.7 mg/L (banded killifish) to 300.6 mg/L (American eel). Chronic exposure studies with fish species exposed to 0.1 mg/L 2,4-D were conducted for 10 months, and resulted in no observable physiological symptoms, however no histopathological investigation was performed in these studies and no detailed methods were available to assess the validity of these studies. Breeding experiments were conducted with the guppies, with no observable effects for those exposed to 2,4-D at 0.1 mg/L for 10 months. This data is not considered valid for use in a WoE for potential endocrine effects since, among other deficiencies; the majority of fish species used in the studies were field collected from the Hudson River and presumably may have been exposed to multiple environmental contaminants. In addition, there were no experimental details provided in the publication, thus the validity of these studies could not be adequately assessed (Klimisch 4).

Avian

Somers *et al.*, 1974

Aqueous solutions of 2,4-D (amine form; no purity information given) at recommended (2.8 - 3.4 kg/ha) and 20x field application rates (44.8 kg/ha) were sprayed on fertilized hen's eggs (*Gallus domesticus*; single comb while leghorn chickens) prior to incubation. (The number of eggs tested in the study was not reported.) Mortalities *in ovo*, at pip, and at hatch were assessed in addition to weight gain of the chicks 3-4 weeks subsequent to hatch. Vent sexing was performed on all chicks that successfully hatched. No adverse effects on any of the measured parameters were noted with respect to 2,4-D application. Residue analysis indicated that internal contents of the egg showed entry of 2,4-D to be far less than proportionate. This study is not relevant because of the unusual method of application, demonstrated low 2,4-D absorption and the lack of information on the nature of the test substance.

Somers *et al.*, 1978a

Hens and cockerels (*Gallus domesticus*) that came from fertile eggs, were subjected to spray contamination with a "10x recommendation dosage" of 2,4-D (commercial formulation "Esteron 99"; purity not reported), and were evaluated for various aspects of reproduction including egg production, egg weight, shell porosity and strength, sperm counts, testes weights, gross appearance of the testes, embryo mortality, hatching success, and chick mortality and malformations, and weight gain. Two incubation experiments were conducted: the first evaluated reproductive success of the offspring as a function of parental spray treatment, the second involved an egg re-treatment with spray contamination restricted to the pre-incubation period. In general, there was no definitive evidence that 2,4-D had any significant effects on chicken reproduction through one generation and into the second. This study is not relevant for the weight of the evidence evaluation because of the lack of dosimetry information and the lack of information on the test substance.

Somers *et al.*, 1978b

Eggs were obtained daily from a commercial strain of Single Comb White Leghorn hens (*Gallus domesticus*). A total of 432 eggs over 12 replicates per treatment for each of the 3 stages of embryonic development were used in the study design. Commercially available formulation of 2,4-D (Trade name: Esteron 99, purity not reported) was applied via spray to the fertilized eggs at a rate of 10x the "recommended rate (11.2 kg/ha)." Late dead germs, pipped dead germs and all chicks were examined for gross abnormalities. Live performances of a subset of successful hatches were assessed until males were 20 weeks of age and females were 16 weeks of age. During this time body weights and mortality were assessed. (Numbers of subjects and experimental design were not reported in detail.) A statistical evaluation of all data associated with incubation failed to disclose a significant interaction involving spray treatment and the stage of embryonic development at application. Furthermore, 2,4-D application had no effect on hatching success, weight gain, or mortality even up to 16-20 weeks of age. This study is not relevant for the weight of the evidence evaluation because of the lack of information on the test substance and uncertain penetration of the test material through the shell. It suggests, however, that there should be little concern for bird eggs inadvertently exposed to formulated 2,4-D.

Mixed Species

Relyea, 2005

Pesticide treatments were replicated six times in 1000 L outdoor microcosms containing algal and 25 animal species including dragonflies, cladocerans, copepods, snails, salamanders, four species of anuran amphibian tadpoles, and various other water bugs. The commercial form of 2,4-D (44.5% active ingredient) was added to microcosms at the manufacturer's recommended application rate (0.117 mL/m2 for 2,4-D). After 13 days of exposure, the microcosms were examined for community diversity, survival, and biomass. Species richness, biomass, and survival of the exposed animal species were not significantly different between control and 2,4-D microcosms. Furthermore, the survival of four tadpole species (Leopard frog, American toad, Gray tree frog, Spring peeper) were unaffected by 2,4-D application. The other insecticides and herbicides evaluated in this test system had marked effects on species richness. There is no specific endpoints related to endocrine activity in this microcosm study, and it is not included in the WoE. The results indicate however, that 2,4-D, when applied appropriately, has no effect on community species richness or tadpole survival.

Supplementary Appendix VII

Mammalian Toxicological Studies with Klimisch Score of 3 or 4[6]

Developmental Studies

Bage *et al.*, 1973

In this study, pregnant NMRI mice were dosed subcutaneously with formulated herbicides containing either 2,4,5-T or a mixture of 2,4,5-T and 2,4-D from days 6-14 of gestation. Doses were 110 mg/kg/day 2,4,5-T; 50 mg/kg/day 2,4,5-T; 110 mg/kg/day 2,4-D/2,4,5-T (2:1); 50 mg/kg/day 2,4-D/2,4,5-T (2:1); and vehicle, for groups A, B, C, D, and E, respectively. The vehicle control group received an equivalent volume of DMSO with 19 mg/mL of the formulated product petroleum distillate mixture. The mice were cesarean-sectioned at gestation day (GD) 18, and the fetuses were evaluated.

Maternal toxicity was not evaluated in this study. There was a marked increase in fetal resorptions, particularly in the high-dose groups. The fetal incidence of cleft palate was also increased in these groups. Decreased mean fetal body weights were also observed in these groups, and the lower body weight fetuses were described as showing retarded skeletal development. The percentage of fetuses with skeletal malformations was also increased, particularly in the 2,4,5-T high-dose group.

This study provides results that are generally considered consistent with other high-dose developmental studies of phenoxy acids at doses that exceed renal clearance of 2,4-D (Garabrant and Philbert, 2002). Because of study design and reporting deficiencies, however, including a route of administration inappropriate for risk assessment, the use of formulated product and absence of a group testing 2,4-D alone, the study is not considered of high quality. This study does not provide any specific information regarding the endocrine-disrupting potential of 2,4-D and is not considered useful for the WoE.

Cavieres *et al.*, 2002

Cavieres *et al.* (2002) tested the effects of a commercial formulation of 2,4-D (7.59%), mecoprop (3.66%), dicamba (0.84 %) and excipients, administered in drinking water, on implantation, litter size, pup weights, and crown rump length inND4 mice. This study combined several studies conducted with two different dosing regimens and smaller dose groups, one from GD 0-15, and one from GD 6-15. Nominal doses of 2,4-D were 0, 0.01, 0.10, 20 and 100 mg/kg/day. Cavieres *et al.* (2002) reported a decrease in live litter size and implantations at all doses of 2,4-D. They also interpreted their data as showing a selective low dose effect of 2,4-D. However, there are a number of significant problems with the study designs and data interpretation that are the bases for these conclusions.

The data were analyzed in depth by Lamb *et al.* (2003), who reviewed the data in the publication in detail, as well as reviewing Cavieres' dissertation, Cavieres (2001) which appeared to serve as the basis

[6] See publication **Table 11** for Klimisch scores and summary comments for published mammalian toxicological studies.

for much of the published information. Lamb *et al.* (2003) submitted a letter to the journal editor with the conclusions from their review. Ashby *et al.* (2003) also submitted an independent critique of the Cavieres *et al.* (2002) conclusions. The major flaws in the Cavieres *et al.* research stem from combining data from multiple experiments, including an experimental design that cannot correctly assess potential effects on implantation, as implantation in mice was completed prior to the start of dosing in several of the studies that were combined by Cavieres *et al.* (2002) for analysis. Cavieres *et al.* (2002) could not have accurately assessed *corpora lutea*, as dams were not sacrificed until weaning of the litters, and *corpora lutea* in rodents will regress during the lactation period. Therefore, the number of potential pups per litter is unknown. Further, Cavieres *et al.* (2002) did not assess litter size immediately after birth, so the possibility that pups were cannibalized and not counted cannot be excluded. Additionally, implantation sites were not counted for all mice. Review of the dissertation further showed data discrepancies between the dissertation and the publication which acted to enhance the statistical significance of the reported effects in some groups of animals. The study authors responded (Cavieres *et al.*, 2003), but do not appear to have addressed all of the critical questions regarding the original publication. In view of testing of a formulation with unknown excipients, the questionable design and inappropriate merging of data from different studies, Cavieres *et al.* (2002) is not included in the WoE.

Collins and Williams, 1971

Syrian golden hamsters were dosed (7-12 litters/group) with test compound from one of three different lots of technical 2,4-D on GD 6-10, inclusive. Doses ranged from 20-100 mg/kg/day. (Six lots of 2,4,5-T were also tested; this phase of the study is not discussed further.) It appears that controls were pooled, so comparison to concurrent control groups is not possible. Dams were caesarean sectioned on gestation day 14, and the uterus examined for resorptions. Fetuses were removed, weighed and examined externally. Approximately 1/3 of fetuses were processed for skeletal evaluation, with the remaining fetuses split into two groups, one for dissection and one for histopathological evaluation.

Slightly decreased fetal viability was observed in the mid- and high-dose compound lot C dams, although no strong dose response was present. There was no treatment-related increase in resorptions or effect on fetal body weight. Percent abnormalities per live litter (according to Table 1) were not statistically significantly higher than the control in any 2,4-D-exposed group. An increased incidence of specific anomalies was reported at doses of 60 mg/kg/day of 2,4-D and higher, although the fetal incidence was low, and dose response was present only for compound lot C. The findings in this study should be considered in the context that hamsters have an extremely high background incidence of malformations, such that they are seldom used in current developmental toxicity evaluations. Although this study has multiple weaknesses that limit its reliability including use of pooled controls and small group size, the results are generally considered consistent with those of other high-dose developmental studies of phenoxy acids at doses that would be expected to exceed renal clearance of 2,4-D. However, the weaknesses in this study are significant.

Dominant Lethal

Oakes *et al.*, 2002a

Male Sprague-Dawley rats (12 weeks old, five/group) were dosed with Tordon 75D® (2,4-D and picloram in formulation) 5 days per week for 9 weeks at doses equivalent to 37.5, 75, and 150 mg/kg/day of 2,4-D, and 9.4, 18.7, and 37.5 mg/kg/day picloram for the low-, mid-, and high-dose groups, respectively. Untreated and positive control (cyclophosphamide [CP]) groups were also used, with a similar exposure regimen. Males were given the opportunity to mate with two untreated females during weeks 2 and 3, 4 and 5, and 8 and 9 of treatment. The males were then left untreated and unmated for 11 weeks, after which they were given the opportunity to mate with four females over a 2-week period. This study was done in two replicates over a 2-year period. In a small additional experiment, control, positive control and high-dose males (2/group) were mated with two untreated females, and the females were allowed to litter to give an F1 generation. At maturity pups from the F1 litters were mated (with offspring of the other female from the same dose group). For both studies, females were killed on gestation day 20, corpora lutea were counted, live and dead fetuses and visible resorptions were counted, and fetuses were weighed and evaluated for visceral and skeletal anomalies. Data were analyzed by analysis of variance (ANOVA) with the Fisher exact test. The litter was considered the unit of analyses for the reproductive studies.

Some rats (dose level unspecified) appeared lethargic shortly after dosing. There was slight body weight loss in the high-dose group through the first 4 weeks, and decreased body weight gain through the remainder of the dosing period. Body weights in the low and mid-dose were equivalent to the controls. There were no dose-related effects on fertility, post-implantation loss, mean litter size or weight, number of runts, incidence of malformations or skeletal variations in breeding results from any mating interval attributable to Tordon 75D®. Pre-implantation loss was increased in the high-dose group for both the 2- to 3-week and 4- to 5-week mating interval litters, (15.2% and 16.2%, respectively, compared to a range of 5.1%-11% in the various control groups); these increases are not statistically significantly different from control. Although the study authors consider these to be possibly treatment-related findings, we consider this unlikely. The incidence of pre-implantation loss is based on comparison of the total of fetuses and resorptions to the number of *corpora lutea*, which is inexact both because of the difficulty in detecting early resorptions (particularly if the uteri are not stained to facilitate this, which was not done in this study), and also because the *carpora lutea* count may be inexact. There was a statistically significant increased incidence of hydronephrosis in high-dose fetuses from the 2- to 3-week mating interval litters (5.8% compared to 0-1.9% in the various control groups). It cannot be determined from the data presented whether these are fetal or litter incidences. The authors do not discuss the treatment relationship of this finding. Hydronephrosis/dilated renal pelvis may be a subjective finding that tends usually to have a moderate background incidence in Sprague-Dawley rats. The incidence reported in this study for the Tordon 75D® high-dose level is within the historical control range reported for Sprague-Dawley rat fetuses reported by Middle Atlantic Reproduction and Teratology Association (MARTA), which examined the incidence of malformations in 223 studies conducted from 1989-1992 (reported in *Handbook of Developmental Toxicity*, R. Hood, ed., CRC Press, 1996). Overall,

Tordon 75D® showed minimal treatment-related effects (if any), and none of the findings characteristic of dominant lethal or male mediated toxicity.

In contrast, the positive control (CP) dosed rats showed increased pre-implantation loss (16.9% and 19.8% for the 2- to 3-week and 8- to 9-week mating interval litters, respectively; an increased number of female fetuses, for the 4- to 5-week mating interval; increased post-implantation losses at all three mating intervals (18.5%-60.7%) and correspondingly decreased litter sizes; and increased malformations for the 4- to 5-week mating interval litters. These results are as would be expected for a positive dominant lethal response.

In the small scale study breeding F1 animals, there were no apparent differences from control in the F2 generation fetuses for either the high-dose Tordon 75D® or the CP groups. The latter study is clearly limited by the very small group size evaluated. The authors conclude that these studies indicate "within the limitations of the power of the study, that herbicide mixtures of 2,4-D and picloram are unlikely to be male reproductive toxicants." The major problems in this study are that 2,4-D was tested only as part of a formulation mixture, and the group size tested was relatively small.

Male Reproductive Toxicity

Blakley, *et al.*, 1989

In this study, male CD-1 mice were dosed via drinking water for 60 days with Tordon 202c and mated to untreated females. Calculated compound intake were 5, 10, and 20 mg/kg/day picloram and 84, 168, and 336 mg/kg/day 2,4-D for the low, medium, and high concentration groups, respectively. Reduced water consumption and high mortality was seen in the adult males at the highest concentration. Females were cesarean sectioned at gestation day 18 and fetuses evaluated. The numbers of implantations, live or dead fetuses, and resorptions were not affected. The investigators note that fetal weight was decreased at the high dose, and that there were increased variations and/or malformations at all doses tested. No pattern of specific malformations was identifiable, and no dose response was present. Variations were reported primarily as the decreased ossification typical of developmental delay, correlating with the decreased mean fetal weight noted in the high-dose group. There are several deficiencies in the study design and reporting that make the results difficult to interpret with respect to developmental toxicity of 2,4-D. First, 2,4-D was not tested separately, but only in conjunction with picloram and with other formulation excipients (not identified). Second, the group size was small (5-14 litters). Third, fetal and litter incidence of specific types of malformations and/or variations were not reported. The study findings are considered inconclusive and not attributable to endocrine disruption. Although it is possible that some toxicity to sperm occurred in the high-dose group, it should be noted that this group was dehydrated and showed high mortality. This study is not included in the WoE evaluation.

Stoker *et al.*, 2007

Stoker *et al.* reported an evaluation of the effects of 2,4-D on pubertal development and thyroid function in the juvenile male Wistar rat. This study was reported in an abstract only, with limited supplementary information presented in a book chapter by Stoker and Zorrilla, 2010; information important for interpreting the results such as impact on body weight of the dosed animals was not reported. The rats were exposed to 0, 3, 30, 100 or 200 mg/kg/day of 2,4-D by oral gavage from postnatal day (PND) 23 to 53, with the two high dose levels markedly exceeding the TSRC. Preputial separation (PPS) was evaluated beginning on PND 33. The rats were necropsied on PND 53 and tissue weights recorded. Serum was collected at necropsy and analyzed for thyroid, testicular and pituitary hormone concentrations. PPS was significantly delayed for 2.7 days in the 200 mg/kg/day group compared to controls. However, no effect on PPS was observed at 100 mg/kg/day. The mean weight of the ventral prostate and levator ani plus bulbocavernosus muscles (LABC) was reported to be significantly decreased in the 200 mg/kg/day group. Testosterone and androstenedione were reported to be significantly decreased following exposure to the highest dose of 2,4-D, but luteinizing hormone and prolactin were not altered at either dose. Triiodothyronine (T3) and thyroxine (T4) were significantly decreased at both 100 and 200 mg/kg of 2,4-D, but not at 30 mg/kg/day. TSH was not evaluated, and no histopathological evaluation of the thyroids was done, so the biological significance of this finding cannot be assessed, but it is not inconsistent with findings in subchronic studies at similar doses exceeding the TSRC. Information on this study is limited, the purity of the test material is not defined, and key information such as body weight effects are not reported. Unfortunately limited details are available regarding the conduct of this study, which otherwise supports that thyroid and male reproductive hormone changes are only evident at doses exceeding the TSRC in rats and are hence not relevant for human risk assessment.

Kim *et al.*, 2002

This study used the Hershberger assay to evaluate potential androgenicity of 2,4-D (50 mg/kg/day) and its metabolite, 2-4-dichlorophenol (DCP; 100 mg/kg/day) in castrated CD rats. Male Sprague-Dawley rats castrated on postnatal day (PND) 42 were treated with 2,4-D or DCP from PND 57-66. Consistent with the OECD and EPA test guidelines, there were 6 animals per treatment group. The authors reported statistically significant increases in ventral prostate, Cowper's gland and glans penis weights with both 2,4-D and DCP treatment, suggesting androgenicity. Furthermore, the authors reported that these compounds potentiated increases in accessory sex tissue weights by testosterone.

It is difficult to fully review the manuscript by Kim *et al.* (2002) as the publication is written in Korean. Still, a few observations on study procedures and results are possible. First, it is unclear what dose of testosterone was used in this study (i.e., the abstract states "1 g/kg", whereas the Hershberger assay description states "1 ug/kg" and Figure 2 on Hershberger body weights states "1 mg/kg/day"). The Hershberger test guidelines (OPPTS 890.1400; OECD 441) recommend doses of 0.2-0.4 mg/kg/day testosterone propionate for the anti-androgenic assessment. In the Kim study, only a single dose level

of 2,4-D was examined (50 mg/kg/day, exceeding the TSRC), whereas a minimum of 2 dose levels are required in the test guidelines to assess androgenicity, and 3 dose levels to assess anti-androgenicity.

The relative potency of 2,4-D relative to testosterone reported in the Kim study also poses questions. 2,4-D was reported to significantly increase seminal vesicle and Cowper's gland weights and interestingly, the magnitude of increases in accessory sex gland weights with 2,4-D were reported to be similar to the increases seen with testosterone treatment. Similarly, DCP at 100 mg/kg/day also were reported to have had effects similar in magnitude to testosterone. A later publication indicates "In a previous study, we showed that these compounds exhibited synergistic androgenic effects by co-treatment with testosterone in the Hershberger assay." The problem is that the testosterone supplemented assessment in the Hershberger is designed to measure potential anti-androgenic activity, not androgenic activity, and it never has been validated for the latter purpose.

On closer inspection, the organ weights reported by Kim *et al.* (2002) differ from other studies using testosterone-supplemented male Sprague-Dawley rats castrated at ~6 weeks of age and with similar terminal body weights. For example, control organ weights in the Kim *et al.* study were 7.66 mg ventral prostate; 65.83 mg seminal vesicle; 151.18 mg levator ani-bulbocavernosus muscle; 29.47 mg glans penis and 2.28 mg Cowper's glands; whereas mean control organ weights in the validation study by Yamasaki *et al.* (2003) were 18.0 mg ventral prostate (range 12.6-21.1); 45.4 mg seminal vesicles (range 39.7-52.5); 209.1 mg levator ani-bulbocavernosus muscle (192.3-236.3); 54.0 mg glans penis (range 48.9-62.6) and 6.9 mg Cowper's glands (range 6.5-8.1). Using the same test system (CD rats castrated at 6 weeks of age), Kang *et al.* (2004) reported accessory sex tissue weights similar to Yamasaki *et al.* (2003).

It also seems unlikely that 2,4-D could have similar androgenic potency to testosterone and not be detected in other androgen receptor binding assays or in *in vivo* studies including androgen-sensitive endpoints. For example, 2,4-D failed to show evidence of androgenic activity in the EOGRT study (Marty *et al.*, 2010), wherein male reproductive organ weights were not increased in parental or F1 males, age at puberty onset was not decreased, anogenital distance was not altered, and sperm parameters and reproductive organ histopathology were unaffected. Because only a single dose group was used by Kim *et al.* it is not possible to evaluate a dose response. Because of these flaws and the lack of consistency of reported findings with other 2,4-D studies this study is excluded from the WoE.

This publication also described an *in vitro* experiment with 2,4-D. It appears these data were also published (in English) in Kim *et al.*, 2005.

Oakes *et al.*, 2002b

Male SD rats (16 weeks old, 5/group) were dosed with Tordon 75D® (2,4-D and picloram in formulation) at 0.125 ml/kg/day, 0.25 ml/kg/day or 0.5 ml/kg 5 days per week for 9 weeks (2,4-D doses of 37.5, 75 and 150 mg/kg/day), and then held unexposed for a 21-week recovery period. Untreated and surfactant control groups were also used, with a similar exposure regimen. At the end of the recovery period,

testes and epididymides were weighed for all groups. Additionally for the high dose group and the control groups, blood was sampled for plasma testosterone levels, and the testes were examined histopathologically. In a second study to evaluate the testicular findings, groups of high dose animals were evaluated after 1, 2, 4 and 9 weeks of treatment for plasma testosterone levels, testes and epididymides weights and testicular histopathology (findings compared to 9-week control groups).

High dose rats appeared lethargic shortly after dosing. There was slight body weight loss in the high dose group at the two-week interval; and it appeared there may have been slightly decreased weight gain during the remaining seven weeks of exposure (not statistically significant compared to controls); also see concluding comment re method deficiencies.

In the second (high dose) study, there were no treatment-related effects on plasma testosterone levels at any interval. Testicular absolute and relative weights were statistically significantly decreased compared to the controls in the high dose group at 9 weeks only. A duration-related increase in incidence and severity of testicular damage was evident histopathologically, characterized by germ cell depletion and shrunken seminiferous tubules. Generally Leydig cells appeared to be conserved. A single high dose rat at 4 weeks showed normal testis weight and histopathological findings in one testis, but enlargement of the other testis with both germinal cell depletion and Leydig cell hypoplasia. The relationship to treatment of the findings in this animal was unclear. One of 5 rats showed germ cell depletion 21 weeks post treatment, so reversibility of the high dose testicular effects was not established. However, similar findings (but unilateral) were made in one surfactant-exposed control rat (at 9 weeks), and one other surfactant exposed animal also showed a mixture of normal and depleted tubules. Certain formulation excipients may therefore have contributed to the observed findings.

There were no treatment-related changes in testicular weights in the mid and low dose groups (evaluated 21 weeks post exposure only); other parameters were not evaluated for these groups. The findings in this study cannot be clearly attributed to 2,4-D alone. The effects seen at the high dose are suggestive of direct cytotoxicity to the germinal cell epithelium rather than endocrine-related changes (as Leydig cells were generally preserved and there were no discernable effects on plasma testosterone levels, although the evaluation of the latter was limited by the small group size). The effects were characterized only at a very high dose of 2,4-D (150 mg/kg/day) in the Tordon®75 formulation, which would be a dose expected to exceed the renal clearance threshold for 2,4-D.

The study quality was generally poor. The description of experimental design in this study is difficult to follow. The group size is relatively small (5/group), which is considered insufficient to reliably evaluate testosterone. The methods do not state whether statistical comparisons were made with the untreated or the surfactant controls. (Recovery group comparisons were made to the untreated controls.)

The high dose study design appeared deficient in that there were no concurrent controls for testosterone, testes weights, and testicular histopathology for the 1, 2, and 4 week intervals. Further, the reporting of body weight data is incomplete (as control weight gain is reported for the 9-week

interval only) so that treatment-related effects on body weight gain cannot clearly be evaluated for the high dose study.

Body weight effects, if any existed, cannot be evaluated at all for the low and mid dose groups from the data presented (as weight change is shown only following the 21-week recovery period for these dose groups). The study design did not include histopathological evaluation of the testes for the low and mid dose groups, and did not evaluate the testes for these groups at all at 9 weeks (when the findings in the high dose pathogenesis study appeared to be the most severe). Therefore, it is not possible establish a NOAEL for this study.

Importantly 2,4-D was not tested in isolation; and it is unknown whether picloram or the other formulation components contributed to the observed effect.

Studies of Female Reproductive Health, Lactation, Nursing Behavior

Stürtz *et al.*, 2010

The authors evaluated the effects on lactation and the control thereof after oral dietary exposure to 2,4-dichlorophenoxyacetic acid (2,4-D) in lactating Wistar rats. The authors reported decreased pup weight gain in all treated groups (2.5, 5, 15, 25, 50, and 70 mg 2,4-D/kg bw) during the first 16 days *postpartum* which they contend is due to interference with hormonal control of lactation at the level of the hypothalamo-pituitary axis. (Note that the dosing units are assumed based on other publications by the same group of authors; these are not clearly specified in the paper.) There are, however, several limitations in the study's design and documentation that weaken their conclusions.

First, it is not clear how the doses to the animals that received 2,4-D in their feed were calculated/verified. Diets were formulated in a manner corresponding to that described below in Stürtz *et al.*, 2006. Apparently individually dosed feed was prepared for each animal every second day by spraying the feed with a solution of 2,4-D. Homogeneity was not reported. It was not reported if the feed was available *ad lib*, or how or when feed consumption was measured. It was stated that analytical verification of target concentrations was conducted, but no results are shown. It was not indicated how excursions from target doses were handled.

Second, it was not indicated that pups were kept away from the feed. Most rat pups begin to eat some amount of chow beginning around PND 13 or 14 (Tyl *et al.*, 2008). This could seriously confound the interpretation of the results. Third, no bedding or nesting materials are mentioned. Female rats about to litter exhibit nesting behaviors and stress to the dams could result from unavailability of bedding or nesting materials. Fourth, because the dams are the exposed individuals, the litter should have been the experimental unit for statistical evaluations. A mean of pup body weights should have been calculated for each litter and these litter means (N=8) should have been the N used for evaluation of pup body weights. Ideally male and female pup weights should be analyzed separately. The tables in the paper appear to present means of individual pup body weights (N = ~64) which will inflate erroneously the power of the study and could lead to incorrect statements of statistical significance. Further,

although litters were culled to result in approximately equal sex ratios, any skewing because of a shortage of males or females could influence the group mean weight.

Fifth, the thrust of this paper appears to be that there are important changes that occur in pups when there are no grossly observable alterations in dams. Nevertheless, the authors note that because there were no differences in general condition, body weight gain, or mortality, the naïve and vehicle control groups were combined. This belies the premise of the study and serves to artificially increase the power of the study by increasing the size of the control group.

Sixth, it is not clear how (or if) the tissues of the anterior pituitary or arcuate nucleus were pooled for analyses related to monoamines and NOS. A table of wet weights for these tissues was not presented, although the authors claim the weights of the tissues were not substantially different. Using wet tissue weights for very small tissue samples is also a potential source of error as small tissue samples are affected greatly by amounts of hydration/dehydration that can occur prior to weighing.

Lastly, the authors evaluated milk production by taking total litter weights before and after previously-starved litters were allowed to suckle for 15 minutes. This is a creative idea but there limited control data and no historical control data for comparison. In particular, the authors' major point is made in an experiment that compared milk production in control animals to those exposed to 50 mg 2,4-D/kg ip and to those exposed to 2,4-D/kg ip plus 100mU oxytocin. The results are dramatic, but they are inconclusive because oxytocin is a strong signal for release of milk from the mammae. A more complete picture would have been gained had there been a fourth group (control plus 100mU oxytocin) to evaluate the impact of oxytocin on non-exposed mammae.

Based on data generated from their procedures mentioned above, the authors reported compound-related changes in pup weight gain, milk production, and hypothalamic concentration of several neurochemicals which they use in constructing a complex argument to explain a potential effect of 2,4-D on milk production in rats. Due to the lack of rigor in obtaining and analyzing the data, their theoretical construct remains speculative at best. One interesting point to note is that there were marked effects on pup body weight, and similar to the Sturtz *et al.* study a lack of clinical signs of toxicity, in the range-finding study of 2,4-D conducted by Shakil *et al.*, 2008 but only at far higher dietary dose levels; supporting the questions regarding diet preparation and the actual doses administered in the Sturtz *et al.* study.

Stürtz *et al.*, 2008

Stürtz *et al.* (2008) exposed lactating Wistar dams to dietary 2,4-D (15, 25, or 50 mg/kg body weight/day) from lactation day (LD) 1 to 7 and maternal behaviors were assessed. Maternal nest building was not altered at any dose of 2,4-D; however, the authors reported altered dam-pup interactions, particularly with respect to nursing. Retrieval, crouching and licking of pups were absent or reduced during the 10-minute monitoring period in 2,4-D-exposed dams. Latency to retrieval and latency for maternal crouching for pup nursing was increased with 2,4-D treatment. The authors also reported that dams consumed more food during the light phase and had high self-grooming. 2,4-D-

treated dams were reported to have increased catecholamine levels and decreased indolamine levels in the arcuate nucleus. Serum prolactin levels were decreased 62-70% in dams given 2,4-D.

One major limitation of this study is that the authors did not appear to use the litter as the unit of analysis.

Rats were dosed via the same procedures described in the Stürtz *et al.* (2006) study; thus, the same issues with the dosing regimen apply. The authors state that dietary concentrations were analytically verified, but these results are not presented, nor do the authors present data on feed consumption or the frequency and formulas used for adjusting dietary concentrations during lactation. Homogeneity was not reported. It was not indicated how excursions from target doses were handled. The dose levels used in this study were high such that the mid and high dose dams would have exhibited nonlinear toxicokinetics (Saghir *et al.*, 2008a), making these data irrelevant for 2,4-D risk assessment.

For maternal care-giving behaviors, dams were tested for maternal behaviors in their home cages; thus, there is no indication that rats were counterbalanced across test sessions or cage locations. The monitoring period for maternal behaviors was only 10 minutes; thus, the significance of altered maternal behaviors over such a short period of time is questionable, particularly when there were no effects on pup mortality or pup body weight gains. For example, the authors report an absence of maternal anogenital licking on PND 5 and 7 in 2,4-D-treated dams; however, it is not possible for these dams to have discontinued anogenital licking in pups for a prolonged period as this stimulation is needed during the first 3 weeks of life for pups to initiate urination and defecation (Brouette-Lahlou *et al.*, 1999). Without maternal anogenital licking, pups would die, which was not reported. Further, pups suffering from maternal neglect are usually not found nested, rather pups are isolated, and often cold and unresponsive. These types of observations were not reported.

Data from vehicle-treated control dams were combined with data from untreated dams, which serves to artificially increase the power of the study by increasing the size of the control group. Prolactin levels are highly variable and subject to change by numerous factors, including stress (cage transport, euthanasia procedures, diurnal variation, etc.). It is unclear if euthanasia was stratified across dose groups, which also may contribute to prolactin effects. Based on inconsistencies in the dosing regimen, high exposure levels, and the lack of corresponding pup effects to support limited monitoring of maternal care, the Stürtz *et al.* study is considered limited and the findings are difficult to interpret.

Stürtz *et al.* (2006)

Stürtz *et al.* (2006) exposed lactating Wistar dams to dietary 2,4-D (15, 25, 50 or 70 mg/kg body weight) from lactation day (LD) 1 to 16 and measured systemic toxicity and toxicokinetic (TK) parameters. 2,4-D did not alter maternal body weight gains, but significantly decreased pup body weight gains at all dose levels. Milk components were analyzed and the authors reported that 2,4-D decreased total lipid content of milk, altered fatty acid content, and resulted in the loss of 3 minor milk proteins. 2,4-D levels in dam serum, dam milk and pup serum also were measured and are discussed below.

One major limitation of this study is that the authors did not use the litter as the unit of analysis.

Furthermore, the Stürtz *et al.* paper lacked detail on critical study procedures. For example, with the limited information presented, the precise dose administered to the animals is difficult to determine. To begin, diet preparations were unusual in the Stürtz *et al.* study as these investigators dissolved 2,4-D in an alcohol solution, sprayed it over the food, and dried it to prepare test diets. The authors state that dietary concentrations were analytically verified, but these results are not presented, nor do the authors present data on feed consumption or the frequency and formulas used for adjusting dietary concentrations during lactation.

The potential issues with the dietary dose administration in this study are further revealed when TK data are examined. The TK data for LD 16 Wistar rats presented in Stürtz *et al.* (2006) can be compared with TK data generated in Saghir *et al.* (2008) using LD 14 CD® rats. The Saghir *et al.* (2008a) study was a GLP study with verified purity of test material, homogeneity, stability and dose concentration analyses. For 2,4-D, marked rat strain differences in TK have not been reported and in fact, Fischer 344 and CD® rats show similar TK with 2,4-D exposures (e.g., Van Ravenzwaay *et al.*, 2003; Saghir *et al.*, 2006; Saghir *et al.*, 2008a). The primary mechanism for 2,4-D clearance is via renal organic acid transporter 1 (OAT-1; Hasegawa *et al.*, 2003), which also is expressed in Wistar rats and shows similar sex steroid regulation across rat strains (Kudo *et al.* 2002, Buist *et al.*, 2002). In a comparison of the Stürtz *et al.* (2006) and Saghir *et al.* (2008a) TK data, marked differences in lactational TK were observed. At 15 mg/kg/day 2,4-D in the diet, Stürtz *et al.* (2006) reported dam serum levels of 26 µg 2,4-D/ml, whereas at 15 mg/kg/day, Saghir *et al.* (2008a) reported dam plasma levels of 3.5 µg 2,4-D/g; thus, Stürtz *et al.* reported 7.4-fold higher dam serum concentrations of 2,4-D with putatively similar dietary exposures (typically, the differential between plasma and serum values for 2,4-D would be minor).

Furthermore, 2,4-D TK exhibited markedly different TK behaviors between these two studies. At 70 mg/kg/day dietary 2,4-D, dam serum levels were 74 µg 2,4-D/ml in the Stürtz *et al.* (2006) study, which represented only 60% of the predicted serum level based on linear TK using serum 2,4-D concentrations at 15 mg/kg/day (26 µg 2,4-D/ml); thus, sublinear increases in serum levels were observed at 70 mg/kg/day. In contrast, Saghir *et al.* (2008a) reported dam plasma levels of 35.3 µg 2,4-D/g at 58 mg/kg/day, which represented 270% of the predicted plasma level based on linear TK using plasma 2,4-D concentrations at 15 mg/kg/day (3.5 µg 2,4-D/g); thus, supralinear increases in plasma levels were observed at 58 mg/kg/day due to saturation of renal clearance in the dams at this dose level. The non-linear TK performance observed in the Saghir *et al.* (2008a) study was consistent with toxicokinetic phenomenon reported previously for 2,4-D (e.g., Gorzinski *et al.*, 1987; Van Ravenzwaay *et al.*, 2003), thus calling into doubt the opposite findings in the Stürtz *et al.* (2006) study.

These TK differences were further reflected in milk levels of 2,4-D, where Stürtz *et al.* (2006) reported milk 2,4-D levels of 29 µg/ml at 15 mg/kg/day, whereas Saghir *et al.* (2008a) reported 2.35 µg 2,4-D/ml at 16 mg/kg/day (milk:serum partitioning of approximately 1 in Stürtz *et al.* compared with milk:plasma partitioning of 0.7 in Saghir *et al.*). At 204 mg/kg/day, the Saghir *et al.* study reported comparable milk 2,4-D concentrations (128 µg/ml) to dams given 74 mg/kg/day 2,4-D in the Stürtz *et al.* study (115

µg/ml). Pup serum levels of 2,4-D also differed (6 µg 2,4-D/ml serum in the Stürtz *et al.* study at 15 mg/kg/day dam dietary dose compared with 2.8 µg 2,4-D/g plasma at 15 mg/kg/day dam dietary dose in Saghir *et al.*). At 70 mg/kg/day, pup serum levels of 2,4-D were less than predicted based on linear TK at 15 mg/kg/day dose in the Stürtz *et al.* study. In the Saghir *et al.* study, pup plasma levels of 2,4-D were markedly increased at doses >57.5 mg/kg/day, again reflecting nonlinear TK and saturation of renal clearance.

Clearly, the TK data indicate the dosing regimens in these two studies differed, which also is reflected in the pup body weight effects. In the Stürtz *et al.* study, pup body weight gains were decreased at all doses of 2,4-D beginning at LD 6 (mean body weight differences of 4-5 g in 2,4-D treated pups by PND 16), whereas in the Saghir *et al.* study, pup body weights were not significantly affected at 100 and 400 ppm (dam doses of 10-15 mg/kg/day and 37-58 mg/kg/day over LD 4-14; dietary concentrations were not adjusted during lactation). Based on differences in dose administration and systemic toxicity between these studies, the Stürtz *et al.* (2006) study is considered limited and the findings are difficult to interpret. Furthermore, the data in the Stürtz *et al.* study do not represent direct endocrine-mediated endpoints, particularly with respect to the activities of concern (estrogen, androgen, thyroid).

Stürtz *et al.*, 2000

Lactating Wistar dams (unclear how many dams/litters were used) were exposed to 50, 70 or 100 mg/kg 2,4-D in DMSO via intraperitoneal (ip) injection. Dams were injected every two days such that pups were exposed to 2,4-D through maternal milk during lactation. Dams were exposed until the offspring reached PND 9 or 15. Pup body weights were significantly decreased at all doses of 2,4-D. Twenty-four hours after the last dose, pups were euthanized; weights of the pups' stomach contents (milk), brain and kidney were decreased with 2,4-D treatment. 2,4-D residue levels were measured in these tissues and blood.

In this study, dams were exposed to 2,4-D by ip injection, which is not a relevant route of exposure. Furthermore, dose levels of 2,4-D were high and likely in the nonlinear TK range. Based on the lack of a relevant route of exposure and the large margin of exposure for 2,4-D (i.e., making findings at nonlinear TK doses irrelevant), these data cannot be used for human risk assessment. On PND 16, 2,4-D-treated pups weighed 40, 65 and 70% of control pup body weights at 100, 70 and 50 mg/kg/day, respectively, indicating that the maximum tolerated dose (MTD) was exceeded at all three dose levels tested. Given the substantial decreases in pup body weights, it is not surprising that some tissue weights (i.e., kidney, stomach contents, brain) differed from the control values. During analytical experiments to determine 2,4-D concentrations in tissues, 2,4-D recovery from spiked samples was only 50-70%. It does not appear as if the litter was the unit of statistical analysis in this study.

Duffard *et al.*, 1995

In an abstract, Duffard *et al.* (1995) described 2,4-D effects on prolactin, progesterone and estrous cycle alterations in F1 generation females only at a dose level (70 mg/kg/day) to dams well above the TSRC and thus irrelevant for risk assessment. Changes were reported in hormone levels for both the control and exposed rats; actual data were not presented but the reporting is considered unreliable. Interestingly the abstract was titled developmental neurotoxicity; however, no data relevant to assessment of neurotoxicity were reported. The Duffard *et al.* study is not considered reliable because of the limited data presented (abstract only) and questionable presentation of findings.

Other Studies

Kobal *et al.*, 2000

Wistar rats (8/sex/dose group) were gavaged with an aqueous 2,4-D formulation (purity or excipients unknown) at doses of 11 or 110 mg/kg/day for 10 days. The high dose exceeds the TSRC in rats. Blood samples were taken 3 days before the start of dosing and on days 6 and 13. Serum T3 and T4 were determined by radioimmunoassay. There were clear decreases in serum T4 in the high-dose group male and females and decreased T3 in high-dose males on the 6th day of the experiment. Although the investigators report that statistically significant differences were observed in the low- dose group (increases in females and decreases in males), examination of the figures suggest that these changes were not biologically significant and may represent the inherent variability of the assay parameters. Data were portrayed graphically only. No other thyroid-related parameters (such as thyroid stimulating hormone [TSH], thyroid weights or thyroid histopathology) were assessed in this study, and the biological significance of the findings cannot be assessed. This study is limited by the unknown nature of the test material and the relatively few parameters assessed.

Pochettino *et al.* (2010)

Pochettino *et al.* (2010) studied the effects of prenatal and postnatal exposure to 2,4-D on oxidative stress in ventral prostate, ovary, and mammary tissue in rats by measuring both the levels of reactive oxygen species (ROS) and the activities of antioxidant enzymes (Tables X and X2). Pregnant rats (6 dosed and 6 control) were exposed to dietary doses of 70 mg/kg/day of 2,4-D daily from 16 days of gestation up to postnatal day (PND) 23. Concentrations were adjusted based on the dams most recent body weight to achieve the approximate mg/kg/day dose, however doses to pups, whose direct feed intake is approximately 2x-3x the dams between PND 17 and 23 would have been exposed to much higher doses. Neither 2,4-D purity nor the specific source of 2,4-D was identified in the study. Dose analyses were either not conducted or the results not reported. 70 mg/kg/day far exceeds the TSRC and has shown significant toxicity in prior rat studies.

F1 young adults (six/sex/dose, reported to be taken from different litters) were sacrificed at PND 45, 60, or 90 at which point ventral prostates, mammary tissue, and ovaries were removed and assayed for ROS levels or antioxidant enzyme activities. To evaluate ROS stress, hydroxyl radical levels, total carbonyl

and thiol levels, and lipid peroxidation (LPO) activity levels were measured (**Table S9**). Antioxidant response was assessed by measuring activities of antioxidant enzymes glutathione s-transferase (GST), catalase (CAT), selenio-glutathione peroxidase (Se-GPx), glutathione reductase (GR), and superoxide dismutases (SODs) (**Table S10**).

2,4-D exposure significantly reduced weight gain in both males and females.

Most of the changes in ROS and anti-oxidant enzyme activity were in the 1-2.5 fold range compared to control. Similar fold differences were seen between control groups at different ages for many of the parameters measured, although the variability in the controls was not statistically analyzed or discussed. Although limited by the single high dose administered and not considered a high quality study, this study provides some evidence of oxidative stress which may explain some of the high dose findings above the TSRC, not relevant to hazard or risk assessment, that otherwise might suggest anti-androgenic activity or adverse effects on steroidogenesis.

Subchronic toxicity in Mammals other than Rodents

Obidike *et al.* (2012)

Obidike *et al.* studied the effects of a 2,4-D formulation in male West African Dwarf goats. Goats were treated by gavage with 75 mg/kg, 100 mg/kg, or 125 mg/kg of 720 g/L 2,4-D (unspecified purity) every 72 hours for a period of 112 days. Control goats were administered no 2,4-D. On the 112th day, all goats were sacrificed and testes and epididymis were removed to investigate the testicular morphology and gonadal (testicular) and extra-gonadal (epididymal) sperm reserves.

There were no significant differences in combined (left and right) testicular sperm reserve at 75 mg/kg. A 1.96-fold and a 2.36-fold decrease were observed at 100 and 125 mg/kg, respectively. Total epididymal sperm reserve in caput, corpus and cauda segments combined was decreased significantly at 75, 100 and 125 mg/kg (1.68-fold, 2.38-fold, and 2.28-fold, respectively). Authors noted that goats dosed with 2,4-D showed hyperemia and edema of the stroma and decreased Sertoli cells. There was no consistent change in the spermatogenic cell population of seminiferous tubules. This study has some significant limitations. First, it is not possible to tell what the administered dose of 2,4-D was, i.e., was the low dose 75 mg/kg 2,4-D or 75 mg/kg of the 720g/L formulation? Second, the nature of the formulation is not specified and excipients are unknown. Only 5 goats per group were evaluated, which is a limited number for assessment of sperm parameters which are inherently variable. The age variation of the goats was significant (greater than 5 months); it is unknown whether younger goats may have differences in testicular structure and function from older goats but this seems likely based on extrapolation from other species. Goats came from different sources; although a single variety of goats was tested there is no assurance the control group was adequately representative. Further, goats were treated during the study for endo- and ecto-parasites; the potential male reproductive toxicity associated with these treatments was not discussed. Little is known about saturation kinetics in goats or other ruminants; based on analogous testicular findings in rats at doses greater than 100 mg/kg/day the

doses in this study are likely to have markedly exceeded the threshold for saturation kinetics. Information in this study is considered too limited to use in the weight-of-the evidence analyses.

Rawlings *et al.*, 1998

In this study, mature female ewes were dosed via gelatin capsules administered directly to the rumen for 43 days (10 mg/kg bw, 3x/week). Serum concentrations of LH, FSH, progesterone, estradiol, thyroxine, cortisol, and insulin were measured, and histopathological examinations were completed of the thyroid, adrenals, pancreas, pituitary, ovaries, oviducts, uterus, and other organs for any treatment-related lesions. 2,4-D exposure resulted in a statistically significant decrease in serum concentrations of thyroxine at 36 days; however, there was no correlating exposure-related effect on thyroid histopathology, and the finding is not considered adverse. Likewise, there was no effect on other hormone concentrations or on the endocrine organs or other organs evaluated. Overall, the evaluation of effects on the thyroid, and on other hormonal concentrations, is too limited to place credence in the reported outcomes. This lack of confidence is based on the limited group size (n=6) and the testing of a single dose level, as well as the unconventional choice of a ruminant species for evaluation, with compound administration directly into the microbial rich rumen. Information in this study is considered too limited to use in the weight-of-the evidence analyses.

References to Appendices (includes references only cited in Appendices; references cited in both Appendices and primary study are referenced in primary study)

Ashby J, Pate I, Tinwell H. (2003) Reported seasonal dependence of herbicide developmental toxicity in mice. [Letter]. Environ Health Perspect. 111:A450-A451.

Bélanger B, Bélanger A, Labrie F, Dupont A, Cusan L, Monfette G. (1989) Comparison of residual C-19 steroids in plasma and prostatic tissue of human, rat and guinea pig after castration: Unique importance of extratesticular androgens in men. J Steroid Biochem. 32(5):695–698.

Behnam-Rassoli M, Herbert LC, Howard V, Pharoah PO, Stanisstreet M. (1991) Effect of propylthiouracil treatment during prenatal and early postnatal development on the neocortex of rat pups. Neuroendocrin. 53(4):321-7.

Bon E, Barbe U, Nunez Rodriguez J, Cuisset B, Pelissero C, Sumpter JP, LeMenn F. (1997) Plasma vitellogenin levels during the annual reproductive cycle of the female Rainbow Trout (*Oncorhynchus mykiss*): establishment and validation of an ELISA. Comp Biochem Physiol. 117B, 1:75–84.

Brouette-Lahlou I, Godinot F, Vernet-Maury E. (1999) The mother rat's vomeronasal organ is involved in detection of dodecyl propionate, the pup's preputial gland pheromone. Physiol Behav. 66:427-236.

Buist SC, Cherrington NJ, Choudhuri S, Hartley DP, Klaassen CD. (2002) Gender-specific and developmental influences on the expression of rat organic anion transporters. J Pharm Exp Ther. 301(1):145-51.

Cavieres MF, Jaeger J, Porter W. (2003) Developmental effects of herbicides in mice. [Letter] Environ Health Perspect. 111(14):A74B.

Cavieres MF. (2001) Reproductive and developmental toxicity of a commercial herbicide formulation in mice. [PhD Thesis]. Madison, WI: University of Wisconsin-Madison.

Dantzler WH, Wright SH. (2003) The molecular and cellular physiology of basolateral organic anion transport in mammalian renal tubules. Biochim et Biophys Acta - Biomembranes 1618, 2, 185–193.

Gaido KW, Leonard LS, Lovell S, Gould JC, Babai D, Portier CJ and McDonnell DP. (1997) Evaluation of chemicals with endocrine modulating activity in a yeast-based steroid hormone receptor gene transcription assay. Toxicol Appl Pharmacol. 143:205-212.

Hanley Jr TR, Watanabe, PG. (1985) Measurement of solid feed consumption patterns in neonatal rats by 141Ce-radiolabeled microspheres. Toxicol Appl Pharmacol. 77:496-500.

Kang IH, Kim HS, Shin JH, Kim TS, Moon HJ, Kim IY, Choi KS, Kil KS, Park YI, Dong MS, Han SY. (2004) Comparison of anti-androgenic activity of flutamide, vinclozolin, procymidone, linuron, and p, p'-DDE in rodent 10-day Hershberger assay. Toxicol. 199:145-159.

Kudo N, Katakura M, Sato Y, Kawashima Y. (2002) Sex hormone-regulated renal transport of perfluorooctanoic acid. Chem Biol Interact. 139:301-316.

Lamb JC, Neal BH, Ginevan ME, Bus JS, Mahlburg WM. (2003) Herbicide effects on embryo implantation and litter size. [Letter re Cavieres et al.]. Environ Health Perspect. 111:A450.

Larsen PR, Ingbar SH. (1992) The thyroid gland. Williams Textbook on Endocrinology. (Wilson, J.D. and Foster, D.W., Eds.) W.B. Saunders Company, Philadelphia, Pennsylvania, pp. 357-487.

Lavado-Autric R, Auso E, Garcia-Velasco JV, del Carmen Arufe M, Escobar del Rey F, Berbel P, Morreale de Escobar G. (2003) Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny. J. Clin. Invest. 111:1073-1082.

Li Z, Zhang H, Gibson M, Li J. (2012) An evaluation on combination effects of phenolic endocrine disruptors by estrogen receptor binding assay. Toxicol In Vitro. 26(6):769-74.

Ma Y, Han J, Guo Y, Lam PK, Wu RS, Giesy JP, Zhang X, Zhou B. (2012) Disruption of endocrine function in in vitro H29SR cell-based and in in vivo assay in zebrafish by 2,4-dichlorophenol. Aquat Toxicol. 106-107: 173-81.

Mak P, McDonnell DP, Weigel NL, Schrader WT, O'Malley BW. (1989) Expression of functional chicken oviduct progesterone receptors in yeast (Saccharomyces cerevisiae). J Biol Chem. 265:21613–21618.

O'Connor JC, Cook JC, Marty MS, Davis LG, Kaplan AM, Carney EW. (2002) Evaluation of Tier I screening approaches for detecting endocrine-active compounds (EACs). Crit Rev Toxicol. 32:521-549.

Palmer SJ, Krueger HO. (1997) 2,4-D (2,4-dichlorophenoxy acetic acid): A 96 hour static acute toxicity test with the leopard frog tadpoles (Rana pipiens). Unpublished study by Wildlife International, Ltd., Easton MD. Project No. 467A-102.

Pham TA, Elliston JF, Nawaz Z, McDonnell DP, Tsai MJ, O'Malley BW. (1991) Antiestrogen can establish nonproductive receptor complexes and alter chromatin structure at target enhancers. Proc Natl Acad Sci. 88:3125-3129.

Pham TA, Hwung YP, Santiso-Mere D, McDonnell DP, O'Malley BW. (1992) Ligand-dependent and -independent function of the transactivation regions of the human estrogen receptor in yeast. Mol Endocrinol. 6:1043-1050.

Rehm S. (2000) Spontaneous testicular lesions in purpose-bred beagle dogs. Toxicol Pathol. 28(6):782-7.

Routledge EJ, Sumpter JP. (1996) Estrogenic activity of surfactants and some of their degradation products assesses using a recombinant yeast screen. Environ Toxicol Chem. 15:241–248.

Sandberg JA, Duhart HM, Lipe G, Binienda Z, Slikker W Jr, Kim CS. (1996) Distribution of 2,4-dichlorophenoxyacetic acid (2,4-D) in maternal and fetal rabbits. J Toxicol Environ Health. 49(5):497-509.

Schlenk D. (2008) Letter to the Editor: Response to: Comments on "Evaluation of Estrogenic Activities of Aquatic Herbicides and Surfactants Using a Rainbow Trout Vitellogenin Assay." Toxicol Sci. 104(1):231–233.

Schoonover CM, Seibel MM, Jolson DM, Stack MJ, Rahman RJ, Jones SA, Mariash CN, Anderson GW. (2004) Thyroid hormone regulates oligodendrocyte accumulation in developing rat brain white matter tracts. Endocrin. 145:5013-5020.

Shibutani M, Woo GH, Fujimoto H, Saegusa Y, Takahashi M, Inoue K, Hirose M, Nishikawa A. (2009) Assessment of developmental effects of hypothyroidism in rats from in utero and lactation exposure to anti-thyroid agents. Reprod Toxicol. 28:297-307.

Sretarugsa P, Wallace RA. (1997) The developing Xenopus oocyte specifies the type of gonadotropin-stimulated steroidogenesis performed by its associated follicle cells. Dev Growth Differ. 39(1):87-97.

Tyl RW, Crofton K, Moretto A, Moser V, Sheets LP, Sobotka TJ. (2008) Identification and interpretation of developmental neurotoxicity effects: A report from the ILSI Research Foundation/Risk Science Institute expert working group on neurodevelopmental endpoints. Neurotoxicol Teratol. 30:349-381.

Yamasaki K, Sawaki M, Ohta R, Okuda H, Katayama S, Yamada T, Ohta T, Kosaka T, Owens W. (2003) OECD validation of the Hershberger assay in Japan: Phase 2 dose response of methyltestosterone, vinclozolin, and p,p'-DDE. Environ Health Perspect. 111:1912-1919.

Zysk JR, Johnson B, Ozenberger BA, Bungham B, Gorski J. (1995) Selective uptake of estrogenic compounds in Saccharomyces cerevisiae: A mechanism for antiestrogen resistance in yeast expressing the mammalian estrogen receptor. In: Blair RM, Fang H, Branham WS, Haas BS, Dial SL, Moland CL, Tong W, Shi L, Perkins R, Sheehan DM. (2000) The estrogen receptor relative binding affinities of 188 natural and xenochemicals: Structural diversity of ligands. Toxicol Sci. 54:138-153.

Supplementary Tables

Table S1. Percent Specific Binding of [^{3}H]-17β-Estradiol to the ER in Presence of 2,4-D (%)

2,4-D Conc. (M)	Run 1	Run 2	Run 3	Mean ± SEM[a]
10^{-11}	99.0	98.8	103.5	100.5 ± 1.59
10^{-10}	95.4	99.1	103.6	99.4 ± 2.4
10^{-9}	97.7	98.1	105.0	100.3 ± 2.4
10^{-8}	99.0	102.0	103.5	101.5 ± 1.3
10^{-7}	99.6	102.9	105.9	102.8 ± 1.8
10^{-6}	99.5	102.7	104.0	102.0 ± 1.3
10^{-5}	99.8	100.1	106.1	102.0 ± 2.0
10^{-4}	101.1	98.4	108.7	102.7 ± 3.1

a. SEM Standard error of the mean.

Table S2. Percent Specific Binding of [^{3}H]-R1881 to the AR in Presence of 2,4-D (%)[a]

2,4-D Conc. (M)	Run 1	Run 2	Run 3	Mean ± SEM
10^{-11}	98.79	98.24	99.89	99.0 ± 0.5
10^{-10}	102.44	99.72	99.55	100.6 ± 0.9
10^{-9}	103.53	97.07	99.95	100.2 ± 1.9
10^{-8}	103.22	99.22	104.25	102.2 ± 1.5
10^{-7}	101.49	99.04	103.98	101.5 ± 1.4
10^{-6}	103.17	98.32	104.51	102.0 ± 1.9
10^{-5}	102.33	97.65	105.58	101.9 ± 2.3
10^{-4}	105.64	99.8	106.32	103.9 ± 2.1

SEM Standard error of the mean.

Table S3. Mean (±SD) Hormone Concentrations Following Treatment with 2,4-D for 48 Hours in the Steroidogenesis Assay

Nominal Concentration (µM)	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3	Mean	± SD	Statistical Significance
	Testosterone (pg/ml)			Fold Difference					
DMSO	842.0	735.0	778.0	1.0	1.0	1.0	1.0	0.0	---
0.0001	776.0	691.7	746.3	0.9	0.9	1.0	0.9	0.0	None
0.001	805.7	696.0	752.0	1.0	0.9	1.0	1.0	0.0	None
0.01	737.7	696.3	726.7	0.9	0.9	0.9	0.9	0.0	None
0.1	752.0	647.0	745.3	0.9	0.9	1.0	0.9	0.0	None
1	796.7	665.0	734.0	0.9	0.9	0.9	0.9	0.0	None
10	787.7	684.3	747.0	0.9	0.9	1.0	0.9	0.0	None
100	714.3	610.3	728.0	0.8	0.8	0.9	0.9	0.1	None
	Estradiol (pg/ml)			Fold Difference					
DMSO	29.8	35.7	37.0	1.0	1.0	1.0	1.0	0.0	---
0.0001	24.7	37.2	36.0	0.8	1.0	1.0	0.9	0.1	None
0.001	24.7	35.7	34.2	0.8	1.0	0.9	0.9	0.1	None
0.01	25.6	34.8	37.8	0.9	1.0	1.0	1.0	0.1	None
0.1	26.5	36.4	34.9	0.9	1.0	0.9	1.0	0.1	None
1	25.3	33.7	35.1	0.8	0.9	0.9	0.9	0.1	None
10	27.0	33.8	37.9	0.9	0.9	1.0	1.0	0.1	None
100	34.6	44.3	45.1	1.2	1.2	1.2	1.2	0.0	Yes, all 3 trials

Table S4. 2,4-D AMA Study Design

Nominal concentration, mg a.e./L	Mean measured concentration, mg a.e./L	Number of replicates	Total number exposed (Stage 51 at initiation)	Number examined at Day 7[a]	Number examined at Day 21[a]	Number examined for thyroid histopath at Day 21
Negative control	<LOQ	4	80	20	60	20
0.4	0.273	4	80	20	60	20
4.0	3.24	4	80	20	59	20
40	38.0	4	80	20	60	20
100	113	4	80	20	60	20

a Examinations included NF developmental stage, hind-limb length, snout-to-vent length, and body weight.

Table S5. Study Design 2,4-D Fish Short Term Reproductive Toxicity Assay

Nominal concentration mg a.e./L	Mean measured concentration mg a.e./L	Number of replicates	Number males per replicate	Number females per replicate	Total number adult fish exposed
Negative control	<LOQ	4	2	4	24
0.4	0.245	4	2	4	24
4.0	3.14	4	2	4	24
40	34.0	4	2	4	24
100	96.5	4	2	4	24

Table S6. P1 Dose Groups for F1-Extended One-Generation Dietary Toxicity Study

Dietary Concentration (ppm)		Targeted Approximate Doses (mg/kg/day)		P No. Rats/sex/dose	Satellite Group Targeted No. Pregnant Females/Dose
Male	Female	Male	Female		
Control	Control	0	0	27	12
100	100	5	5	27	12
300	300	15	15	27	12
800	600	40	30	27	12

Table S7. 2,4-D Dietary Concentration Adjustments During Lactation and Post-weaning

Exposure Period	TMI Increase[a]	Unadjusted Concentration (ppm)	Adjustment Factor	Adjusted Concentration (ppm)
LD 7-14	3.1X	100, 300, 600	2	50, 150, 300
LD 14-21	3.8X	100, 300, 600	3	33, 100, 200
PND 21-28[b]	2.4X	100, 300, 600	2	50, 150, 300
PND 28-35[c]	1.9X	F: 100, 300, 600 M: 100, 300, 800	2	F: 50, 150, 300 M: 50, 150, 400

a Relative to non-pregnant adult females (from Saghir et al., 2008a).

b TMI = test material intake based on feed consumption data derived from male CRL:CD(SD) rats at PND 23-28 from Marty *et al.*, 2003.

c After PND 35, pups returned to unadjusted dietary concentrations.

Table S8. Body weight and Testicular Histopathology in the Subchronic Toxicity Study in Dogs with 2,4-D. Schulze, 1990 (Study No. 2184-115)

Dose (mg/kg/day)	Animal Number	Individual Body Weight (kg)		Histopathology Testes	Histopathology Prostate
		Week 0	Week 13		
Control	27494	6.5	10.2	-	NE
	27495	6.9	9.8	-	NE
	27496	5.6	8.3	-	NE
	27497	8.2	12.9	-	NE
	27498	6.5	10.2	-	NE
0.3	27504	6.3	8.1	-	NE
	27505	8.0	10.5	-	NE
	27506	5.8	7.8	-	NE
	27507	7.6	10.1	-	NE
	27508	7.3	11.0	-	NE
1.0	27514	6.4	9.0	-	NE
	27515	5.9	9.0	-	NE
	27516	7.7	-[1]	-	NE
	27517	8.2	10.8	-	NE
	27518	7.5	11.0	-	NE
3.0	27524	7.6	9.2	-	NE
	27525	5.7	7.7	Unremarkable "small" at necropsy	NE
	27526	6.4	7.8	-	NE
	27527	7.4	9.6	-	NE
	27528	6.1	8.6	-	NE
10.0	27534	6.7	*	Hypospermatogenesis-moderate; Giant cells-moderate	NE
	27535	6.0	8.1	Hypospermatogenesis-moderate; Giant cells-moderate	NE
	27536	6.6	8.5	Hypospermatogenesis-slight; Giant cells-minimal	NE
	27537	8.1	7.9	-	NE
	27538	5.8	*	Unremarkable "small" at necropsy	NE

*Dead animal

NE Not evaluated

[1] Number illegible in report copy

Table S8 (Cont.). Body Weights and Testicular and Prostate Histopathology in the Second 13-Week Dietary Toxicity Study of 2,4-D in Dogs. (Dalgard, 1993a) (Lab Project ID: HWA 2184-125)

Dose (mg/kg/day)	Animal Number	Individual Body Weight (kg)		Histopathology testes	Histopathology prostate
		Week 1	Week 14		
Control	G29625	7.7	10.7	Hypospermatogenesis/juvenile testis-slight; Giant cells-minimal	
	G29626	7.9	10.9	-	Inactive/ juvenile prostate
	G29627	6.8	9.4	Giant cells-minimal	
	G29628	9.4	12.2	Giant cells-minimal	
0.5	G29633	9.0	11.4	-	Inactive/ juvenile prostate
	G29634	7.1	9.2	Giant cells-minimal	
	G29635	7.4	11.3	-	
	G29636	9.8	11.4	Giant cells-minimal	
1.0	G29641	9.3	12.8	Giant cells-minimal	
	G29642	7.8	11.4	-	
	G29643	6.8	10.1	-	Inactive/ juvenile prostate; inflammation
	G29644	8.3	10.1	Giant cells-minimal	
3.75	G29649	9.8	11.9	Giant cells-minimal	
	G29650	10.2	11.8	-	
	G29651	7.9	9.9	Hypospermatogenesis/juvenile testis-slight; Giant cells-slight	Inactive juvenile prostate
	G29652	8.7	8.9	-	
10/7.5	G29657	7.0	9.3	Giant cells-minimal	Inactive/Juvenile prostate
	G29658	8.2	9.9	Hypospermatogenesis/juvenile testis-minimal; Giant cells-minimal	Inactive/juvenile prostate
	G29659	8.2	9.4	Hypospermatogenesis/juvenile testis-slight; Giant cells-minimal	Inactive /juvenile prostate
	G29660	6.7	8.1	-	Inactive juvenile prostate

*Dead animal

Table S9. Changes in ROS levels upon 2,4-D exposure (70 mg/kg/day) (Pochettino *et al.* 2010)

	Hydroxyl radical			Carbonyl			Thiol			LPO		
	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90
Ventral prostate	↑	↑	↑	↑	↑	↑	-	-	-	↑	↑	↑
Ovary	-	-	↑	↑	-	-	-	-	-	↑	↑	↑
Mammary	-	-	-	-	-	-	-	↓	↓	↑	↑	↑

Table S10. Changes in anti-oxidant enzyme activity upon 2,4-D exposure (70 mg/kg/day) (Pochettino *et al.* 2010)

	GST activity			CAT activity			Se-GPx activity			GR activity		
	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90	PND 45	PND 60	PND 90
Ventral prostate	↑	↑	↑	↑	↑	↑	↑	↑	-	-	-	↓
Ovary	↓	↓	-	-	↓	-	↑	↑	↑	-	-	↓
Mammary	↓	↓	↓	-	↓	↓	↓	↓	↓	-	-	-

a No changes were observed in SOD activity at any interval in the ventral prostate, ovary or breast.